

Warsaw University of Life Sciences – SGGW
Institute of Economics and Finance
Department of Econometrics and Statistics

**QUANTITATIVE METHODS
IN ECONOMICS**

**METODY ILOŚCIOWE W BADANIACH
EKONOMICZNYCH**

Volume XXVII, No. 3

Warsaw 2026

EDITORIAL BOARD

Editor-in-Chief: Bolesław Borkowski
Deputy Editor-in-Chief: Hanna Dudek
Managing Editors: Michał Gostkowski, Grzegorz Koszela
Theme Editors:
Econometrics: Bolesław Borkowski
Multivariate Data Analysis: Wiesław Szczesny
Mathematical Economics: Zbigniew Binderman
Data Science: Michał Gostkowski
Financial Engineering: Monika Krawiec
Labor Market Analysis: Joanna Landmesser-Rusek
Statistical Editor: Wojciech Zieliński
Technical Editors: Jolanta Kotlarska, Elżbieta Saganowska
Language Editor: Agata Cienkusz
Native Speaker: Yochanan Shachmurove
Managing Editorial Assistant: Luiza Ochnio

SCIENTIFIC BOARD

Adnene Ajimi (University of Sousse, Tunisia)
Heni Boubaker (University of Sousse, Tunisia)
Peter Friedrich (University of Tartu, Estonia)
Paolo Gajo (University of Florence, Italy)
Agnieszka Gehringer (University of Göttingen, Germany)
Anna Maria Gil-Lafuente (University of Barcelona, Spain)
Jaime Gil-Lafuente (University of Barcelona, Spain)
Vasile Glavan (Moldova State University, Moldova)
Francesca Greselin (University of Milano-Bicocca, Italy)
Ana Kapaj (Agriculture University of Tirana, Albania)
Jirawan Kitchaicharoen (Chiang Mai University, Thailand)
Yuriy Kondratenko (Black Sea State University, Ukraine)
Vassilis Kostoglou (Alexander Technological Educational Institute of Thessaloniki, Greece)
Karol Kukula (University of Agriculture in Krakow, Poland)
Kesra Nermend (University of Szczecin, Poland)
Nikolas N. Olenov (Russian Academy of Sciences, Russia)
Alexander N. Prokopenya (Brest State Technical University, Belarus)
Yochanan Shachmurove (The City College of The City University of New York, USA)
Mirbulat B. Sikhov (al-Farabi Kazakh National University, Kazakhstan)
Marina Z. Solesvik (Nord University, Norway)
Ewa Syczewska (SGH Warsaw School of Economics, Poland)
Achille Vernizzi (University of Milan, Italy)
Andrzej Wiatrak (University of Warsaw, Poland)
Dorota Witkowska (University of Lodz, Poland)

ISSN 2082-792X
e-ISSN 2543-8565

Department of Econometrics and Statistics, WULS – SGGW
(Katedra Ekonometrii i Statystyki SGGW w Warszawie)

Warsaw 2026, Volume XXVII, No. 3
Journal homepage: <https://qme.sggw.edu.pl>



Warsaw University of Life Sciences Press, Nowoursynowska 161, 02-787 Warsaw
Tel. 22 593 55 23

e-mail: wydawnictwo@sggw.edu.pl
wydawnictwo@sggw.edu.pl



Wydawnictwo SGGW



wydawnictwosggw

CONTENTS

Michał Thlon – Energy Demand as a Moderator of the Energy Transition in Africa: A Two-Way Fixed Effects Panel Analysis for 27 Countries, 2000–2022	123
Jakub Krawczyk, Michał Wielechowski – Short-Term WIG20 Volatility around Interest-Rate Decision Announcements by the Monetary Policy Council of the National Bank of Poland: Evidence from 2005–2025	132
Jarosław Gorący – Article 33a of the VAT Act - Improved Financial Liquidity for Importers in Poland and the Risk of Abuse	143
Ewa Wasilewska – Alfabetyzm statystyczny a pułapki myślenia. Analiza luki kompetencyjnej na styku edukacji średniej i wyższej	166
Kaniz Sultana Chy, Md. Shoriful Islam Chowdhury, Md Nurul Islam Chowdhury, Md Ashiqul Islam, Md Asiful Islam – Mitigating Class Imbalance in U.S. Financial Security: A Latent Feature Augmentation Approach for High- Purity Fraud Detection	183

ENERGY DEMAND AS A MODERATOR OF THE ENERGY TRANSITION IN AFRICA: A TWO-WAY FIXED EFFECTS PANEL ANALYSIS FOR 27 COUNTRIES, 2000–2022

Michał Thlon  <https://orcid.org/0000-0001-9627-7773>

Department of Applied Economics
Krakow University of Economics, Poland
e-mail: thlonm@uek.krakow.pl

Abstract: Analysing a balanced panel of 27 African economies over 2000–2022, I document a “renewable share paradox”: in 63% of countries, absolute renewable capacity grew while its share in total installed capacity declined from 54.9% to 46.3%. Using a two-way fixed effects estimator, I find a robust negative within-country relationship between log GDP per capita and the renewable capacity share (β ranging from -10.9 to -21.7 pp), supporting the demand-squeeze hypothesis. Targeted public investment in renewables can partially offset this effect.

Keywords: two-way fixed effects, energy transition, renewable energy share, demand squeeze, Sub-Saharan Africa, panel data

JEL classification: C23, Q42, Q43, O13, O55

INTRODUCTION

Standard approaches to monitoring the energy transition rely predominantly on absolute metrics: total installed renewable capacity in megawatts, annual investment volumes, and electricity generation quantities. These indicators are routinely reported in international frameworks including the SDG7 tracking system and national Renewable Energy Action Plans. However, absolute metrics can generate a misleading signal: a country may simultaneously record growth in absolute renewable capacity and a decline in the renewable share of its electricity system, if non-renewable capacity grows faster. Figure 1 documents this pattern for 27 African economies over 2000–2022, showing that 17 of 27 countries (63%) experienced a decline in the renewable capacity share despite widespread absolute capacity growth.

<https://doi.org/10.22630/MIBE.2026.27.3.9>



Sub-Saharan Africa provides a particularly important laboratory for this dynamic. The region combines the world's lowest electricity access rates - approximately 600 million people without reliable supply - with among the fastest rates of demographic and economic growth globally. Population growth of 2.5–3% annually, combined with rapid urbanisation and expanding industrial activity, implies a structural acceleration of electricity demand that will persist for decades. At the same time, many African countries enter this growth trajectory with electricity systems still dominated by hydropower, leaving them exposed to a specific form of share erosion when new fossil capacity is added to meet incremental demand.

The existing literature on Africa's energy transition has concentrated primarily on barriers to renewable investment - financing constraints, regulatory risk, grid infrastructure deficits - and on the effectiveness of specific policy instruments such as feed-in tariffs and competitive auctions [Peters et al. 2024; Schwerhoff and Sy 2019]. Far less attention has been paid to the structural role of demand growth itself as a moderator of relative transition progress. This paper addresses that gap.

This paper makes three contributions. First, it introduces the demand-squeeze hypothesis as a structural explanation for the renewable share paradox: in 63% of the 27 countries studied, absolute renewable capacity grew while the share declined, a pattern invisible to monitoring frameworks based on absolute metrics [Akrofi 2021; Mossisa and Han 2025]. Second, it applies a two-way fixed effects (TWFE) estimator - absorbing all time-invariant country heterogeneity and common year-level shocks - to isolate the within-country association between GDP per capita growth and the renewable energy share, yielding a robust negative coefficient across five specifications and two dependent variables. Third, it identifies the asymmetric role of hydropower legacy: countries where hydro exceeded the sample median share of total capacity in 2000 face a structural amplification of the demand-squeeze, as any fossil addition immediately erodes a high renewable baseline. The paper is organised as follows. Section 2 describes the data and empirical methodology. Section 3 presents and interprets the results. Section 4 concludes.

LITERATURE REVIEW

The present paper addresses a gap at the intersection of these two strands. Whereas the barriers literature asks what prevents more renewable capacity from being added, this paper asks what determines whether the capacity that is added is sufficient to maintain - let alone improve - the renewable share as total electricity demand expands. The demand-squeeze hypothesis we test is structurally distinct from the access-oriented framing of Dagnachew et al. [2020]: the concern here is not with countries that lack electricity, but with countries where economic growth generates new demand at a pace that outstrips even successful renewable deployment. Identifying this channel within countries over time, net of all country-level and year-level confounders, is what the two-way fixed effects estimator delivers.

A second strand of scholarship examines Africa's electricity systems through the lenses of energy diversification, access, and spatial spillovers. Akrofi [2021] analyses diversification trends across the continent and finds that aggregate statistics obscure substantial heterogeneity in national trajectories. Mossisa and Han [2025] examine the structural tensions between electricity access expansion and emissions intensity in rapidly urbanising Sub-Saharan economies. Ndayishimiye et al. [2025] document the political-economic drivers of generation overcapacity in selected Sub-Saharan countries, pointing to the role of fossil-fuel incumbents in shaping capacity additions. Jeetoo [2022] employs a spatial panel estimator to examine spillovers in renewable energy consumption across the region. Dagnachew et al. [2020] integrate demand-side projections with supply-side policy scenarios to model electricity access pathways under alternative climate targets. This literature advances the understanding of Africa's energy landscape considerably, but does not model demand growth as a source of renewable share erosion or estimate the within-country relationship between income and renewable share using panel variation.

Research on Africa's energy transition has concentrated predominantly on the supply side. A substantial body of work documents the barriers that slow renewable investment: financing constraints and the high risk premiums attached to sovereign debt in Sub-Saharan economies, grid infrastructure deficits that raise the cost and complexity of connecting variable renewables, and regulatory environments that have historically favoured incumbent fossil-fuel utilities [Peters et al. 2024; Schwerhoff and Sy 2019; Deichmann et al. 2011]. The policy instrument most carefully evaluated in this literature is competitive procurement; South Africa's Renewable Energy Independent Power Producer Procurement Programme, launched in 2011, is widely regarded as the most successful implementation on the continent, having mobilised substantial private investment and added several gigawatts of wind and solar capacity by the early 2020s [Peters et al. 2024]. This body of work yields important design lessons for renewable energy policy, but it treats demand growth as a background condition rather than as an active moderator of whether the transition succeeds or fails in relative terms.

DATA AND METHODOLOGY

Data

The analysis uses three publicly available data sources. Installed electricity capacity by technology is sourced from the IRENA Renewable Energy Statistics database [IRENA 2025], providing annual MW data for 2000–2022. GDP per capita in constant 2015 USD is from the World Bank World Development Indicators [World Bank 2026]. Public renewable energy investment in millions of 2022 USD is from the IRENA Public Finance Database [IRENA 2025].

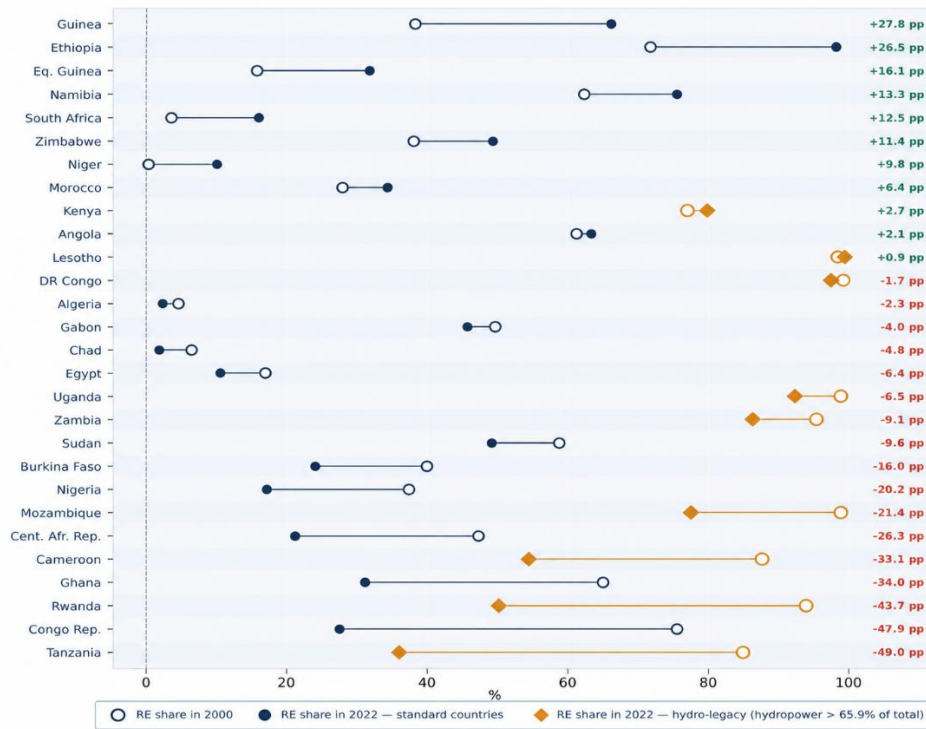
The country sample comprises 27 African economies for which complete data are available for the full 2000–2022 period, including both Sub-Saharan and North African countries (Algeria, Egypt, Morocco, Sudan among the latter). The balanced panel yields 621 country-year observations. The primary dependent variable is the renewable energy capacity share - total renewable installed capacity as a percentage of total installed capacity. Table 1 presents descriptive statistics including the newly added maximum column.

Table 1. Descriptive statistics

Variable	N	Mean	Median	SD	Min	Max	2000	2022
Panel A: Outcome variables								
RE capacity share (%)	621	49.8	46.8	30.0	0.1	100.0	54.9	46.3
RE generation share (%)	621	56.2	52.0	31.5	0.0	100.0	65.0	52.0
Panel B: Key regressors								
GDP per capita (const. 2015 USD)	621	2,283	1,163	2,897	228	13,073	1,534	2,624
Public RE investment (USD mn)	621	92.3	19.5	194.7	0.0	1,825	2.0	150.9

Source: IRENA (2025), World Bank (2026). Panel of 27 African economies, 2000–2022
(N = 621 country-year observations)

Figure 1. Renewable Energy Capacity Share Change, 2000–2022 (27 African Economies)



Source: IRENA (2025); World Bank WDI (2026). Open circle = 2000 value; filled marker = 2022 value. Gold diamonds = hydro-legacy countries (hydropower > 65.9% of total installed capacity in 2000). Change in RE share shown in rightmost column.

Empirical specification

The baseline specification is a two-way fixed effects model:

$$REshare_{it} = \alpha_i + \gamma_t + \beta \ln(GDPpc_{it}) + \varepsilon_{it}$$

where α_i are country fixed effects and γ_t are year fixed effects. Standard errors are clustered at the country level. The coefficient β captures the within-country association between log GDP per capita and the renewable share, holding constant all time-invariant characteristics and common macroeconomic shocks.

The baseline specification is intentionally parsimonious, following the standard approach in within-country panel energy studies [Jeetoo 2022; Dagnachew et al. 2020]. Country fixed effects absorb time-invariant confounders - resource endowments, geography, institutional quality - that drive cross-country differences in renewable shares, while year fixed effects control for global technology cost trends, oil price cycles, and international climate agreements. The within-country

coefficient on log GDP per capita therefore captures the demand-squeeze channel net of these confounders.

Five specifications are estimated: (1) the baseline TWFE model described above; (2) adds log public RE investment and interacts log GDP per capita with a hydropower legacy indicator - a binary variable equal to one for countries where hydropower exceeded 65.9% (sample median) of total installed capacity in the base year 2000, capturing the greater vulnerability of hydro-dominated systems to share erosion when any fossil capacity is added; (3) excludes South Africa and Egypt to test sensitivity to the two largest economies; (4) replaces the dependent variable with renewable generation share; and (5) uses a first-difference estimator with changes in log GDP and log public investment as regressors.

Establishing a fully causal interpretation of the demand-squeeze coefficient would ideally require an instrument that shifts within-country income growth without directly affecting the renewable share. Plausible candidates face a non-trivial exclusion restriction problem in this context: rainfall shocks affect hydro generation directly, and commodity price windfalls alter the political economy of fuel subsidies in ways that independently influence capacity mix decisions [Staiger and Stock 1997]. In the absence of a credible instrument for a 27-country African panel, the two-way fixed effects design represents the strongest available identification strategy. The consistency of the negative coefficient across all five specifications provides reassurance that the result is not an artefact of any single modelling choice.

RESULTS

Table 2 presents all five specifications. The coefficient on log GDP per capita is negative in all specifications and statistically significant at the 10% level in four of five. In the baseline specification (column 1), a 10% within-country increase in GDP per capita is associated with approximately a 1.3 percentage point decline in the renewable capacity share ($\beta \times \ln(1.10) = -13.72 \times 0.0953 \approx -1.3$ pp), holding constant all time-invariant country characteristics and common year trends. This magnitude is economically meaningful: across the 17 countries where the renewable share declined, average GDP per capita grew by 58% over the sample period, implying a demand-driven share reduction of approximately 8 pp - closely matching the observed mean decline of 8.6 pp.

The robustness of the negative sign across all five specifications, including different samples and dependent variables, is the central result: the demand-squeeze mechanism is not artefactual. The larger coefficient in specification (4) using generation share (-21.7 vs -13.7) reflects the compounding effect of thermal plants operating at higher utilisation rates than variable renewables, which accelerates the generation share deterioration relative to the capacity share.

Table 2. Panel regression results: determinants of renewable energy share

Variable	(1) RE cap share	(2) Hydro split	(3) Excl. ZAF/EGY	(4) RE gen share	(5) First diff.
log GDP per capita	-13.72 [†] (7.31)	-10.92* (4.43)	-12.47 [†] (7.13)	-21.66* (10.36)	-
Δlog GDP per capita	-	-	-	-	-1.89 (3.34)
Δlog Public RE investment	-	-	-	-	0.21 [†] (0.11)
Country FE / Year FE	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes	No/Yes
N	621	575	575	621	594
R ²	0.044	0.075	0.047	0.073	0.012

Notes: Dependent variable is RE capacity share (%) in columns (1)–(3), generation share (%) in (4), annual change in capacity share (pp) in (5). Standard errors clustered by country in parentheses. [†]p < 0.10, *p < 0.05.

Source: IRENA (2025); World Bank WDI (2026). Authors' calculations

Ten countries recorded improvements in their renewable capacity share over the period, and their trajectories reveal three distinct escape routes from the demand squeeze. In Ethiopia and Guinea, large-scale hydropower investments - most notably the Grand Ethiopian Renaissance Dam and Guinea's Souapiti project, both delivering first power in the late 2010s - added renewable capacity at a pace that outstripped overall system growth, effectively reversing what had been a fossil-dominated expansion. In South Africa and Morocco, proactive procurement frameworks directed new capacity additions toward renewables: South Africa's REIPPPP programme and Morocco's Noor solar complex channelled private investment into wind and solar at scale, decoupling capacity expansion from fossil-fuel additions. A third group, exemplified by Niger, benefited from a base effect: starting from a near-zero renewable share in 2000, even modest additions of rural solar mini-grids and small wind capacity translated into double-digit percentage-point improvements. These contrasting trajectories reinforce the paper's central argument: the demand squeeze is not inevitable, but overcoming it requires either the scale of public investment or the institutional architecture of targeted procurement - neither of which is costless or automatic.

Column (5) provides evidence from the first-difference estimator. Annual changes in log GDP per capita are not significantly associated with annual RE share changes (-1.89, p = 0.57), consistent with the demand-squeeze operating over multi-year investment horizons. The coefficient on changes in log public RE investment is positive and marginally significant at the 10% level ($\beta = 0.21$, p = 0.06), indicating that years of above-average public renewable investment are associated with modest short-run improvements in the renewable share. This provides qualified support for

the view that targeted, concentrated public finance can partially counteract the structural demand effect.

CONCLUSIONS

This paper demonstrates that the TWFE estimator, applied to a 27-country African panel over 2000–2022, identifies a robust negative within-country relationship between GDP per capita growth and the renewable energy share. Across five specifications using capacity and generation shares, alternative samples, and first-difference estimation, the demand-squeeze mechanism - whereby fossil fuel capacity additions outpace renewable investment in rapidly growing economies - is consistently supported. A 10% within-country GDP increase is associated with approximately a 1.3 pp decline in the renewable capacity share.

Two methodological implications follow. First, energy transition assessments should incorporate relative share metrics alongside absolute capacity indicators; the two can move in opposite directions, with the share carrying greater relevance for welfare outcomes related to carbon intensity and fossil fuel price exposure. Second, the marginal offset provided by public RE investment surges (column 5) supports investment frameworks calibrated explicitly to outpace total capacity growth, not merely to expand absolute renewable megawatts.

The first-difference estimates in Specification (5) allow a tentative quantification of the investment challenge. Average public renewable energy investment in the countries experiencing share declines grew from approximately USD 2 million to USD 151 million per country over the study period - a 75-fold increase. Under the Specification (5) coefficient ($\beta = 0.21$), this rate of investment growth translates into a predicted share improvement of roughly 0.04 percentage points per year, or approximately 0.9 percentage points cumulatively over 22 years. Against the observed mean share decline of 8.6 percentage points in this group, the investment expansion that occurred offset only a modest fraction of the structural demand pressure. The implication is not that public finance is ineffective but that, to overcome the demand squeeze, it must be sustained at a rate calibrated explicitly against total system expansion - a materially more demanding benchmark than the simple growth of renewable megawatts.

The demand-squeeze mechanism is structurally general. Rapidly growing economies in South Asia and parts of Latin America - combining high population growth, rapid urbanisation, and expanding electricity access requirements - may exhibit the same paradox. The TWFE framework developed here provides a replicable methodology for testing this hypothesis in other regional contexts.

REFERENCES

- Akrofi M. M. (2021) An Analysis of Energy Diversification and Transition Trends in Africa. *International Journal of Energy and Water Resources* 5(1), 1-12.
- Dagnachew A. G., Poblete-Cazenave M., Pachauri S., Hof A. F., Van Ruijven B. and Van Vuuren D. P. (2020) Integrating Energy Access, Efficiency and Renewable Energy Policies in Sub-Saharan Africa: a Model-Based Analysis. *Environmental Research Letters*, 15(12), 125010.
- Deichmann U., Meisner C., Murray S., Wheeler D. (2011) The Economics of Renewable Energy Expansion in Rural Sub-Saharan Africa. *Energy Policy*, 39(1), 215-227.
- IRENA (2025) Renewable Energy Statistics 2025. Abu Dhabi: International Renewable Energy Agency.
- Jeetoo J. (2022) Spatial Interaction Effect in Renewable Energy Consumption in Sub-Saharan Africa. *Renewable Energy*, 190, 148-155.
- Mossisa A. T., Han A. T. (2025) Transitioning to Renewable Electric Energy in Rapidly Urbanizing Sub-Saharan Africa: Challenges and Opportunities. *Energy Strategy Reviews*, 61, 101842.
- Ndayishimiye V., Bakkabulindi G., Miyingo E. W., Nibagwire D. (2025) Power Generation Overcapacity in Selected Sub-Saharan African Countries: Political-Economic Drivers and Grid Infrastructure Challenges. *Frontiers in Energy Research*, 13, 1549844.
- Peters R., Berlekamp J., Kabiri C., Kaplin B. A., Tockner K., Zarfl C. (2024) Sustainable Pathways towards Universal Renewable Electricity Access in Africa. *Nature Reviews Earth & Environment*, 5(2), 137-151.
- Schwerhoff G., Sy M. (2019) Developing Africa's Energy Mix. *Climate Policy*, 19(1), 108-124.
- Staiger D. O., Stock J. H. (1997) Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3), 557-586.
- World Bank (2026) World Development Indicators. Washington DC: The World Bank Group. Available at <https://databank.worldbank.org/source/world-development-indicators> (accessed March 2026).

**SHORT-TERM WIG20 VOLATILITY AROUND INTEREST-RATE
DECISION ANNOUNCEMENTS BY THE MONETARY POLICY
COUNCIL OF THE NATIONAL BANK OF POLAND:
EVIDENCE FROM 2005–2025**

Jakub Krawczyk

Faculty of Economics

Warsaw University of Life Sciences – SGGW, Poland

e-mail: s223729@sggw.edu.pl

Michał Wielechowski  <https://orcid.org/0000-0002-1335-8971>

Institute of Economics and Finance

Warsaw University of Life Sciences – SGGW, Poland

e-mail: michal_wielechowski@sggw.edu.pl

Abstract: This study examines short-term WIG20 volatility around 235 interest-rate announcements by the Monetary Policy Council of the National Bank of Poland in 2005-2025. Volatility is measured from daily log returns in three- and five-day windows and analysed separately for decisions involving interest-rate changes and status quo decisions. Descriptive statistics show generally higher pre-announcement volatility and particularly pronounced skewness and kurtosis before interest-rate-change announcements. Wilcoxon signed-rank and Mann-Whitney U tests show that pre-announcement volatility is significantly higher only for decisions involving interest-rate changes in the three-day window. Between-group differences are also significant in the three-day window, but only at the 10% level.

Keywords: monetary policy announcements, interest-rate decisions, short-term WIG20 volatility, Monetary Policy Council of the National Bank of Poland, non-parametric tests

JEL classification: E52, G10, G14

INTRODUCTION

Monetary policy decisions, especially regarding interest rates, are among the most important factors that determine macroeconomic stability and affect financial markets [Mishkin 1995; Rigobon, Sack 2004]. Changes in interest rates affect the cost of capital, market participants' expectations and the valuation of financial assets. Consequently, central bank announcements are closely monitored by financial markets and their participants because they may provide information about current policy conditions and the future macroeconomic outlook. Evidence from interest-rate and foreign-exchange futures [Ederington, Lee 1993] as well as forex markets [Andersen, Bollerslev 1998] shows that scheduled macroeconomic announcements may lead to rapid price adjustments and temporary changes in volatility. Although monetary policy announcements affect equity markets across countries [Nikkinen, Sahlström 2004; Wongswan 2009], their effects may differ depending on the characteristics of the domestic financial market.

Under the efficient market hypothesis, publicly available information should be incorporated into asset prices before an official announcement [Fama 1970]. Therefore, market reactions should primarily reflect the unexpected component of monetary policy decisions. According to this assumption, Bernanke and Kuttner [2005] demonstrate that stock market reactions are mainly driven by unexpected actions in monetary policy, whereas anticipated policy decisions have more limited effect on equity prices. Gürkaynak et al. [2005] further show that monetary policy surprises may influence longer-term expectations, indicating that market participants continuously update their assessments in response to macroeconomic and monetary policy information.

While asset prices react to news, central bank announcements may also trigger market volatility and risk perception. Bomfim [2003] reports an increase in realised stock-market volatility around monetary policy announcements. By contrast, Chen and Clements [2007] and Vähämaa and Äijö [2011] demonstrate that option-implied volatility decreases after central bank meetings. This decline supports the view that market uncertainty is resolved once the policy decision is announced. Lyócsa et al. [2019] identify realised-volatility jumps around scheduled central bank announcements in G7 equity markets. Extending this evidence to emerging markets, Jain [2025] proves that unexpected monetary policy actions generate stronger and more persistent volatility responses because of lower market liquidity and differences in central-bank credibility.

Despite extensive international evidence, research on short-term volatility around monetary policy announcements in the Polish capital market remains limited. Kołodziejczyk [2016] examines WIG20 returns and volatility in response to National Bank of Poland decisions and identifies an asymmetric market reaction depending on the policy outcome. Interest-rate increases are associated with lower volatility, whereas rate cuts increase stock returns without producing a statistically significant

effect on volatility. However, empirical evidence for Poland remains limited, particularly regarding differences in short-term volatility adjustments across monetary policy decision outcomes.

The contribution of the study is that it examines 235 interest-rate decision meetings of the Monetary Policy Council of the National Bank of Poland (MPC) over the 2005-2025 period, providing a long-term perspective on the relationship between monetary policy announcements and WIG20 volatility. Moreover, it distinguishes between rate-change and status quo decisions and compares event-level pre- and post-announcement volatility in three-day and five-day windows. It also tests whether the magnitude of the pre-to-post volatility change differs between the two decision outcomes.

Accordingly, the study assesses whether short-term WIG20 volatility declines after MPC interest-rate announcements and whether this adjustment differs between rate-change and status quo decisions. To verify it, the following research hypotheses are tested:

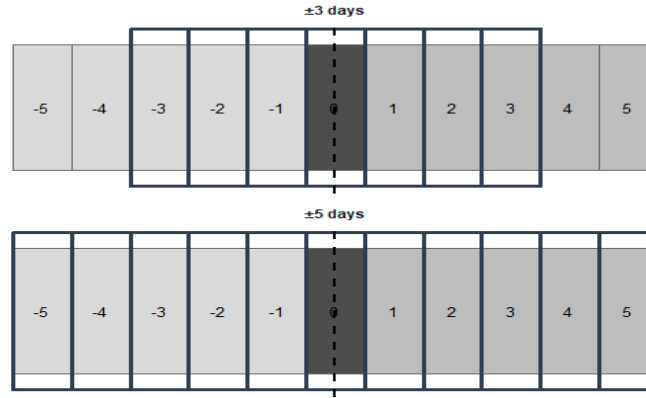
- H1: Short-term WIG20 volatility is higher before than after MPC interest-rate announcements.
- H2: The pre-to-post change in short-term WIG20 volatility differs between rate-change and status quo decisions of the MPC.

METHODOLOGY

The objective of the study is to assess changes in the short-term volatility of the WIG20 index around monetary policy announcements and to determine whether these changes differ between decisions involving an interest-rate adjustment and decisions maintaining the status quo. The empirical investigation covers 235 interest-rate decision meetings of the Monetary Policy Council of the National Bank of Poland (MPC) over the period from January 2005 to December 2025.

To analyse volatility around these events, two pairs of symmetric observation windows surrounding the announcement day are defined, as illustrated in Figure 1. Following the methodological framework of Wielechowski et al. [2026], relatively short event windows are applied to capture the immediate market adjustment around MPC decisions. Short observation intervals limit the potential impact of other macroeconomic factors and political events that may affect stock prices over longer periods.

Figure 1. Timeline of the event windows surrounding the MPC interest rate decision meetings



Source: own elaboration using R

MPC announcement date is designated as event day $t = 0$. For $k \in \{3, 5\}$, the pre-announcement window covers trading days $[-k, -1]$, whereas the corresponding post-announcement window covers trading days $[1, k]$. Event day $t = 0$ is excluded from both calculation periods to maintain a clear temporal separation between pre- and post-announcement observations and to avoid assigning the announcement-day return arbitrarily to either window. As interest rate decisions typically occur during trading hours, daily closing prices on the event day capture a state simultaneously reflecting pre-announcement expectations and the immediate market reaction following the release.

The schedule of MPC decision meetings was obtained from the official website of the National Bank of Poland, while daily closing values of the WIG20 index were retrieved from the Stooq.pl financial database. One meeting was excluded because its event window overlapped with the window of an adjacent meeting. This procedure ensured that the analysed event windows did not share daily quotation dates.

To determine whether market reactions differ depending on the monetary policy outcome, MPC meetings were divided into two mutually exclusive subgroups. The rate-change subgroup includes MPC meetings resulting in an increase or decrease in interest rates, whereas the status quo subgroup consists of meetings at which interest rates remained unchanged. The final sample contains 58 rate-change decisions and 177 status quo decisions.

Daily logarithmic returns of the WIG20 index were calculated as:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \cdot 100\%, \quad (1)$$

where P_t denotes the closing value of the WIG20 index on trading day t , and P_{t-1} represents its closing value on the preceding trading day.

For each meeting i , event-window length k , and period $s \in \{pre, post\}$, market volatility was measured as the sample standard deviation of daily logarithmic returns:

$$\sigma_{i,k}^s = \sqrt{\frac{1}{k-1} \sum_{t \in \mathcal{T}_k^s} (r_{i,t} - \bar{r}_{i,k}^s)^2}, \quad (2)$$

where:

$$\mathcal{T}_k^{pre} = \{-k, \dots, -1\}, \mathcal{T}_k^{post} = \{1, \dots, k\},$$

and $\bar{r}_{i,k}^s$ denotes the mean logarithmic return for meeting i , window length k , and period s . Accordingly, $\sigma_{i,k}^{pre}$ represents volatility before the announcement, whereas $\sigma_{i,k}^{post}$ denotes volatility after the announcement.

The event-level pre-to-post change in volatility was defined as:

$$d_{i,k} = \sigma_{i,k}^{pre} - \sigma_{i,k}^{post}. \quad (3)$$

A positive value of $d_{i,k}$ indicates that volatility was higher before the MPC interest-rate announcement than after it and therefore represents a post-announcement decline in volatility. Conversely, a negative value indicates that volatility increased following the MPC announcement.

The empirical strategy corresponds to the two research hypotheses formulated in the Introduction. Hypothesis H1 assumes that short-term WIG20 volatility is higher before than after MPC interest-rate announcements. It was tested for the overall sample and separately for the rate-change and status quo subgroups.

Because the pre- and post-announcement observations refer to the same MPC meeting, the comparisons were conducted using the right-tailed Wilcoxon signed-rank test for paired samples [Wilcoxon 1945]. For each analysed sample and event-window length, zero differences were excluded, while the absolute non-zero differences $|d_{i,k}|$ were ranked in ascending order. The sum of ranks associated with positive differences was calculated as:

$$W_k^+ = \sum_{i=1}^{n_k} R_{i,k} I(d_{i,k} > 0), \quad (4)$$

where $R_{i,k}$ denotes the rank assigned to $|d_{i,k}|$, $I(\cdot)$ is the indicator function, and n_k is the number of non-zero paired differences. Since the expected direction of the relationship was specified in advance, the right-tailed test was used. The statistical hypotheses can be expressed as:

$$H_{0,1}: \theta_{d,k} \leq 0,$$

against:

$$H_{1,1}: \theta_{d,k} > 0,$$

where $\theta_{d,k}$ represents the location shift of the distribution of paired differences.

A sufficiently high value of W_k^+ provides evidence that volatility is systematically higher before than after the MPC announcement.

Hypothesis H2 assumes that the pre-to-post change in short-term WIG20 volatility differs between rate-change and status quo decisions. To test this hypothesis, the distributions of the event-level differences $d_{i,k}$ were compared between the two independent decision-type subgroups. The comparison was conducted separately for the three-day and five-day window specifications.

A two-sided Mann-Whitney U test for independent samples [Mann, Whitney 1947] was applied. For each event-window length, the values of $d_{i,k}$ from the rate-change and status quo subgroups were combined and ranked. The statistic for the rate-change subgroup was calculated as:

$$U_k = R_{RC,k} - \frac{n_{RC}(n_{RC} + 1)}{2}, \quad (5)$$

where $R_{RC,k}$ is the sum of the ranks assigned to the event-level differences in the rate-change subgroup, n_{RC} is the number of rate-change meetings, and n_{SQ} is the number of status quo meetings. The complementary statistic may be expressed as:

$$U_{SQ,k} = n_{RC}n_{SQ} - U_k. \quad (6)$$

The tested hypotheses were:

$$H_{0,2}: F_{RC,k}(d) = F_{SQ,k}(d),$$

against:

$$H_{1,2}: F_{RC,k}(d) \neq F_{SQ,k}(d),$$

where $F_{RC,k}$ and $F_{SQ,k}$ denote the distributions of the pre-to-post volatility differences in the rate-change and status quo subgroups, respectively.

Basic descriptive statistics were calculated for the pre- and post-announcement volatility measures, including the mean, median, minimum and maximum values. In addition, the third standardised moment was calculated to evaluate skewness, while the fourth standardised moment was used to assess kurtosis.

The normality of the event-level differences $d_{i,k}$ was additionally examined using the Shapiro–Wilk test [Shapiro, Wilk 1965]. The test results, together with the observed skewness and kurtosis, indicated substantial departures from normality and supported the use of non-parametric inference.

The primary significance level was set at $\alpha = 0.05$. For completeness, results were also evaluated at the less restrictive significance level of $\alpha = 0.10$. All empirical calculations and statistical tests were conducted using version 4.1.1 of the R software environment.

RESULTS AND DISCUSSION

To assess the short-term volatility of the WIG20 index around MPC interest-rate announcements, standard deviations of daily logarithmic returns were calculated for the three-day and five-day pre- and post-announcement windows. The results are presented for 235 MPC meetings, including 58 rate-change and 177 status quo decisions.

Table 1 reports the descriptive statistics and distributional characteristics of the volatility measures. In most specifications, both mean and median volatility are higher before than after the MPC interest-rate announcement. In the three-day window, mean volatility declines from 1.253 to 1.155 for the overall sample and from 1.515 to 1.176 for rate-change decisions. The corresponding decline for status quo decisions is smaller, from 1.167 to 1.148. A similar pattern is observed in the five-day window, although the post-announcement median for rate-change decisions slightly exceeds its pre-announcement value.

Table 1. Descriptive statistics and distributional characteristics of WIG20 volatility around MPC interest-rate announcements in 2005-2025

Time window	Decision type	Period	Mean	Median	Min	Max	Skewness	Kurtosis
3 days	Overall	Pre	1.253	1.066	0.059	9.543	3.680	29.133
		Post	1.155	1.009	0.043	4.269	1.552	6.647
	Rate Change	Pre	1.515	1.093	0.206	9.543	3.642	20.631
		Post	1.176	1.036	0.043	3.479	1.052	4.193
	Status Quo	Pre	1.167	1.055	0.059	4.266	1.363	6.396
		Post	1.148	1.009	0.060	4.269	1.726	7.538
5 days	Overall	Pre	1.330	1.145	0.253	6.925	2.695	15.489
		Post	1.252	1.115	0.212	4.205	1.523	6.011
	Rate Change	Pre	1.540	1.149	0.253	6.925	2.342	10.038
		Post	1.372	1.177	0.452	4.205	1.558	6.250
	Status Quo	Pre	1.262	1.145	0.357	4.481	1.605	7.348
		Post	1.213	1.077	0.212	3.605	1.493	5.746

Source: own elaboration and calculation based on NBP and Stooq.pl data using R

All volatility distributions are positively skewed, indicating infrequent observations with particularly high volatility. The strongest skewness is observed before MPC announcements in the overall sample and the rate-change subgroup. In the three-day window, pre-announcement skewness reaches 3.680 and 3.642, respectively, compared with 1.552 and 1.052 after the announcements.

Kurtosis also indicates pronounced leptokurtosis, especially before rate-change meetings. In the three-day window, pre-announcement kurtosis equals 29.133 for the overall sample and 20.631 for rate-change decisions, declining

substantially after the announcement. Thus, the results suggest that periods preceding rate-change MPC meetings are characterised not only by higher average volatility but also by heavier-tailed distributions and a greater tendency for extreme volatility values to occur. An exception is the three-day status quo subgroup, where skewness and kurtosis are higher after the announcement.

The pronounced positive skewness and high kurtosis indicate substantial departures from normality and the presence of extreme volatility observations. As an additional diagnostic check, the Shapiro-Wilk tests rejected the normality of all event-level pre-to-post volatility difference distributions ($p < 0.001$). These distributional characteristics support the application of non-parametric statistical procedures.

Table 2 presents the right-tailed Wilcoxon signed-rank tests used to evaluate H1 and the two-sided Mann-Whitney U tests used to evaluate H2. A positive median difference indicates that volatility is higher before the MPC announcement than after.

Table 2. Wilcoxon signed-rank and Mann-Whitney U test results for changes in WIG20 volatility around MPC announcements in 2005-2025

Event window	Test	Comparison	Effect estimate (median difference)	Test statistic	p-value
3 days	Wilcoxon signed-rank test, one sided	All meetings (pre vs post)	0.072	15423	0.068*
		Rate change (pre vs post)	0.198	1074	0.046**
		Status quo (pre vs post)	0.027	8349	0.245
	Mann-Whitney U test, two-sided	$d_{i,k}$: rate change vs status quo	0.038	5962	0.065*
5 days	Wilcoxon signed-rank test, one sided	All meetings (pre vs post)	0.092	15534	0.055*
		Rate change (pre vs post)	0.099	989	0.152
		Status quo (pre vs post)	0.078	8767	0.096*
	Mann-Whitney U test, two-sided	$d_{i,k}$: rate change vs status quo	0.004	5723	0.190

Note: ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

Source: own elaboration and calculation based on NBP and Stooq.pl data using R

Hypothesis H1 assumes that short-term WIG20 volatility is higher before than after MPC interest-rate announcements. At the primary significance level of $\alpha = 0.05$, H1 receives partial support. A statistically significant reduction in volatility is

identified only for rate-change decisions in the three-day window, where the median pre-to-post difference equals 0.198.

The remaining Wilcoxon results are not significant at the 5% level. At $\alpha = 0.10$, additional evidence emerges. For the overall sample, the differences are significant in both the three-day window, and the five-day window. For status quo decisions, only the five-day result is significant at the 10% level.

The five-day result for rate-change decisions is not significant at either level despite a positive median difference of 0.099. Therefore, support for H1 is not uniform across windows and decision types. At the 5% level, it is restricted to the immediate adjustment following rate-change decisions, while the 10% level provides broader but weaker evidence of declining post-announcement volatility.

Hypothesis H2 assumes that the event-level pre-to-post volatility change differs between rate-change and status quo decisions. The Mann-Whitney U test does not support H2 at the 5% level. In the three-day window, the difference between the subgroups equals 0.038, providing marginal evidence only at the 10% level.

For the five-day window, the difference is smaller, amounting to 0.004, and is not statistically significant. Thus, any asymmetry in the volatility response appears concentrated in the period immediately surrounding the MPC announcement and becomes statistically indistinguishable over the longer horizon.

The findings provide limited evidence that MPC announcements are followed by lower short-term WIG20 volatility. Moreover, results significant at the 10% level should be interpreted as suggestive rather than conclusive evidence. The significant three-day result for rate-change decisions is consistent with the uncertainty-resolution mechanism [Vähämaa, Äijö 2011] and with Kołodziejczyk [2016], who identifies an asymmetric response of the Polish equity market. However, the adjustment is less uniform than in developed markets, where central bank announcements affect volatility across a broader range of meetings [Bomfim 2003; Lyócsa et al. 2019]. The concentration of the strongest result in the three-day window is also not fully consistent with Jain [2025], who reports more persistent effects in emerging markets.

CONCLUSIONS

This study examined short-term changes in WIG20 volatility around 235 interest-rate announcements by the Monetary Policy Council of the National Bank of Poland between 2005 and 2025. The results indicate that the market response varies across decision outcomes and event windows. At the 5% significance level, pre-announcement volatility is higher than post-announcement volatility only for rate-change decisions in the three-day window. At the 10% level, additional, though weaker, evidence is observed for the overall sample in both event windows and for status quo decisions in the five-day window. Differences in pre-to-post volatility changes between rate-change and status quo decisions are not significant at the 5%

level, although the three-day comparison provides marginal evidence at the 10% level.

The findings suggest that the strongest volatility adjustment is concentrated in the period immediately surrounding rate-change announcements. This may be relevant for short-term portfolio risk assessment and hedging decisions. However, the results should be interpreted as associations rather than evidence of a causal stabilising effect of MPC decisions.

The study is limited by its focus on the WIG20 index and by the lack of distinction between expected and unexpected monetary policy decisions. Moreover, rate-change announcements are analysed jointly, without separating interest-rate increases from interest-rate cuts, although these decisions may generate different market reactions. Addressing these limitations constitutes a challenge for future research.

REFERENCES

- Andersen T. G., Bollerslev T. (1998) Deutsche Mark–Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies. *The Journal of Finance*, 53(1), 219-265. <https://doi.org/10.1111/0022-1082.85732>
- Bernanke B. S., Kuttner K.N. (2005) What Explains the Stock Market's Reaction to Federal Reserve Policy? *The Journal of Finance*, 60(3), 1221-1257. <https://doi.org/10.1111/j.1540-6261.2005.00760.x>
- Bomfim A. N. (2003) Pre-Announcement Effects, News Effects, and Volatility: Monetary Policy and the Stock Market. *Journal of Banking & Finance*, 27(1), 133-151. [https://doi.org/10.1016/S0378-4266\(01\)00211-4](https://doi.org/10.1016/S0378-4266(01)00211-4)
- Chen E.-T. (John), Clements A. (2007) S&P 500 Implied Volatility and Monetary Policy Announcements. *Finance Research Letters*, 4(4), 227-232. <https://doi.org/10.1016/j.frl.2007.07.002>
- Ederington L. H., Lee J.H. (1993) How Markets Process Information: News Releases and Volatility. *The Journal of Finance*, 48(4), 1161-1191. <https://doi.org/10.1111/j.1540-6261.1993.tb04750.x>
- Fama E. F. (1970) Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383-417. <https://doi.org/10.2307/2325486>
- Gürkaynak R. S., Sack B., Swanson E. (2005) The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models. *American Economic Review* 95(1), 425-436. <https://doi.org/10.1257/0002828053828446>
- Jain A. (2025) How Do Central Bank Announcements Affect Stock Market Volatility? A Quantitative Event Study. *International Journal For Multidisciplinary Research*, 7(6), 621-632. <https://doi.org/10.36948/ijfmr.2025.v07i06.62132>
- Kołodziejczyk H. Z. (2016) The Impact of Monetary Policy Announcements on Stock Market Index in Poland. *Research Papers in Economics and Finance*, 1(1), 7-16. <https://doi.org/10.18559/ref.2016.1.1>

- Lyócsa Š., Molnár P., Plíhal T. (2019) Central Bank Announcements and Realized Volatility of Stock Markets in G7 Countries. *Journal of International Financial Markets, Institutions and Money*, 58, 117-135. <https://doi.org/10.1016/j.intfin.2018.09.010>
- Mann H. B., Whitney D. R. (1947) On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1), 50-60. <https://doi.org/10.1214/aoms/1177730491>
- Mishkin F.S. (1995) Symposium on the Monetary Transmission Mechanism. *Journal of Economic Perspectives*, 9(4), 3-10. <https://doi.org/10.1257/jep.9.4.3>
- Nikkinen J., Sahlström P. (2004) Scheduled Domestic and US Macroeconomic News and Stock Valuation in Europe. *Journal of Multinational Financial Management* 14(3), 201-215. <https://doi.org/10.1016/j.mulfin.2003.01.001>
- Rigobon R., Sack B. (2004) The Impact of Monetary Policy on Asset Prices. *Journal of Monetary Economics*, 51(8), 1553-1575. <https://doi.org/10.1016/j.jmoneco.2004.02.004>
- Shapiro S. S., Wilk M. B. (1965) An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3-4), 591-611. <https://doi.org/10.1093/biomet/52.3-4.591>
- Vähämaa S., Äijö J. (2011) The Fed's Policy Decisions and Implied Volatility. *Journal of Futures Markets*, 31(10), 995-1010. <https://doi.org/10.1002/fut.20503>
- Wielechowski M., Czech K., Weremczuk A., Próchniak M. (2026) Stock Market Reactions to Donald Trump-Involved U.S. Presidential Elections: Evidence from the S&P 500 in 2016, 2020, and 2024. *International Journal of Management and Economics*, 62(1), 26-37. <https://doi.org/10.2478/ijme-2026-0004>
- Wilcoxon F. (1945) Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6), 80-83. <https://doi.org/10.2307/3001968>
- Wongswan J. (2009) The Response of Global Equity Indexes to U.S. Monetary Policy Announcements. *Journal of International Money and Finance*, 28(2), 344-365. <https://doi.org/10.1016/j.jimonfin.2008.03.003>

ARTICLE 33A OF THE VAT ACT - IMPROVED FINANCIAL LIQUIDITY FOR IMPORTERS IN POLAND AND THE RISK OF ABUSE

Jarosław Gorący  <https://orcid.org/0009-0009-6650-8046>

Doctoral School
University of Szczecin
e-mail: jaroslaw.goracy@phd.usz.edu.pl

Abstract: Article 33a of the Polish VAT Act is the Polish model of postponed VAT accounting, allowing import VAT to be settled in the VAT return instead of being paid at customs clearance. Using aggregated administrative data for 2020–2024 obtained from the Analytical Centre Department of the Tax Administration Chamber in Warsaw, the article examines the scale, commodity structure and concentration of imports accounted for under this mechanism. Descriptive analysis, commodity concentration measures and the Herfindahl-Hirschman Index (HHI) are used to identify changes and patterns that may be relevant to customs and tax risk profiling. The results show that Article 33a covered approximately one quarter of the value of Poland’s imports and was concentrated mainly in industrial, raw-material and technological goods. The findings indicate that aggregated customs data can support the initial identification of unusual trade patterns, although such patterns should be treated as risk signals rather than evidence of abuse.

Keywords: Article 33a of the Polish VAT Act, simplified import procedure, accounting in the tax return, foreign trade, CN 2841

JEL classification: F14, H25, K34

INTRODUCTION

Article 33a of the Polish VAT Act regulates the Polish model of postponed VAT accounting. It allows a VAT-registered taxable person to account for import VAT directly in the VAT return instead of paying it at customs clearance. This improves the importer’s financial liquidity, as it does not require funds to be

<https://doi.org/10.22630/MIBE.2026.27.3.11>



temporarily frozen. At the same time, however, the mechanism shifts a significant part of tax control to the stage of verification of tax returns.

Article 33a does not reduce the VAT liability itself, but changes the time and method of accounting for the tax. Consequently, the mechanism combines a significant liquidity benefit for compliant businesses with a shift from tax collection at customs clearance towards subsequent verification based on tax returns, customs declarations and risk analysis. In this article, imports accounted for under Article 33a are also referred to as imports under the simplified import procedure.

The literature identified for this study concerning Article 33a and comparable postponed VAT accounting mechanisms is predominantly legal, institutional and descriptive. To the author's knowledge, the reviewed literature does not provide a quantitative analysis of the scale, commodity structure and concentration of imports accounted for under Article 33a in Poland. One reason for this may be the limited public availability of sufficiently detailed administrative data. The dataset used in this study was obtained on an individual request from the Analytical Centre Department of the Tax Administration Chamber in Warsaw rather than from a generally available statistical database.

Against this background, the main aim of the study is to examine the scale, commodity structure and concentration of imports accounted for under Article 33a in Poland in 2020–2024 and to assess the usefulness of aggregated customs data for customs and tax risk profiling. The contribution of the study is therefore twofold. First, it provides a quantitative assessment of a mechanism that has so far been discussed mainly from a legal and institutional perspective. Second, it demonstrates how commodity-level and geographical concentration analysis may be used as an initial screening tool for identifying import flows requiring more detailed examination. The study does not seek to identify individual cases of fraud or to estimate the VAT gap.

LITERATURE REVIEW

The literature devoted directly to Article 33a of the VAT Act is limited. The most important study is the article by Krychowski [2023], which provides a concise description of the simplification applicable to the importation of goods, the general rules governing the tax liability, the obligation to pay VAT on imports under Article 33, and the conditions for using Article 33a. The author emphasises the liquidity function of the mechanism, while also pointing out that its application depends on the fulfilment of specific statutory conditions [Krychowski 2023].

In the Polish literature, the subject appears more frequently in commentaries on the VAT Act. Article 33a is discussed, among others, by Bartosiewicz [2026], Michalik [2024] and Matarewicz [2023]. These commentaries present Article 33a as a special method of accounting for tax due on imports, analysed together with Articles 33, 86 and 99 of the VAT Act. The common conclusion of this literature is that the provision has a liquidity and technical function, but its application is

conditional on formal requirements and the correct reporting of the tax in the tax return/JPK file.

A general tax-law perspective is provided in the publications by Mastalski [2021] and Smoleń and Wójtowicz [2021]. These authors do not focus on Article 33a itself, but explain the structure of value added tax, the importation of goods as a taxable transaction, and the relationship between import VAT and the destination principle. Krychowski [2023] adds that the importation of goods is a factual event giving rise to tax liability even where customs duty is suspended, reduced or does not arise.

For the customs-law dimension, the commentary on customs law by Laszuk, Komorowski and Michalski [2022] is relevant. It is useful in explaining customs representation, which under Article 33a serves as an organisational and safeguarding condition. The customs literature also makes it possible to link Article 33a with the Union Customs Code, in particular release for free circulation, simplified declarations, entry in the declarant's records and AEO status.

In the English-language literature, the Polish Article 33a may be compared with solutions referred to as postponed VAT accounting, deferred import VAT accounting or an import VAT reverse charge. These terms are not completely uniform or identical, because individual countries define the conditions for access to the declaration-based accounting of import VAT differently. Nevertheless, the economic effect is common: the taxable person does not pay VAT at customs clearance, but reports output tax and, if the relevant conditions are met, input tax in the VAT return.

In the United Kingdom guidance issued by HMRC (His Majesty's Revenue and Customs), postponed VAT accounting is described directly as a method that allows import VAT to be declared and recovered in the same return, rather than paid in advance on importation and recovered later. The guidance also states that the taxable person must be VAT-registered and that, where goods are imported on behalf of the taxable person, it is important to indicate this method correctly in the import declaration [HMRC 2025].

A similar approach is visible in OECD documents. In the International VAT/GST Guidelines, the organisation presents VAT as a tax of particular importance in the context of growing international trade and points to the need to coordinate VAT rules in cross-border transactions in order to limit the risk of double taxation or unintended non-taxation. Although the OECD Guidelines focus mainly on services and intangibles, their starting point—taxation in the jurisdiction of consumption—also justifies the taxation of imported goods while maintaining neutrality for taxable persons engaged in economic activity [OECD 2017].

Where the tax is not physically collected at the border, tax-return data, ex post control and risk profiling become more important. Alexopoulos et al. show that analysis of VAT transaction networks and machine-learning methods can support the identification of tax risks, and that their approach may be implemented after taxable persons have submitted their returns. Although their study does not deal

directly with imports, it provides a sound methodological justification for analysing large volumes of customs and tax data, commodity concentration and unusual changes in the value of imports under the Article 33a mechanism [Alexopoulos et al. 2025].

The control and risk dimension is developed in the work of Keen and Smith [2007], who describe VAT as an efficient tax that is nevertheless vulnerable to abuse where cross-border settlements, deduction rights and tax refunds are involved. Although their study does not directly concern Article 33a, it supports the approach adopted in this article: analysis of large data sets, commodity concentration, changes in value and unusual import profiles.

Studies concerning the analysis of customs data are also relevant from a similar perspective. Park, Kim and Cha [2022] analyse the possibilities of risk analysis in detecting customs-law irregularities on the basis of import declarations, indicating that knowledge of changing trade flows is important for risk management. Mai et al. [2021], in turn, emphasise the variability over time of certain trade patterns and, consequently, of fraud, which requires an analytical approach to the subsequent selection of declarations for inspection. These studies do not deal directly with VAT or import facilitations, but demonstrate the need for data analysis as part of post-clearance controls aimed at combating potential customs fraud [Park, Kim, Cha 2022; Mai et al. 2021].

Since quantitative analyses depend on the adequate quality of statistical data, attention should also be paid to the reliability of international trade data. Markowicz and Baran [2021] examine the quality of intra-Community trade data for selected CN commodity groups.

The literature review therefore shows that direct studies of Article 33a are scarce and are predominantly legal or descriptive, while quantitative studies of tax and customs risk focus mainly on broader problems of VAT fraud, customs irregularities and data-based risk detection. The present study connects these two strands of literature by applying quantitative analysis to the Polish postponed VAT accounting mechanism itself. In this sense, Article 33a is considered not only as a liquidity facilitation for businesses, but also as an area in which customs and tax data may be used for ex post risk profiling.

LEGAL FRAMEWORK OF ARTICLE 33A OF THE VAT ACT

The direct legal basis for Article 33a is Article 211 of Council Directive 2006/112/EC of 28 November 2006 on the common system of value added tax. This provision leaves Member States free to determine the detailed rules for the payment of VAT on importation and, at the same time, allows them to permit a solution under which VAT due on importation is not paid at the time of importation, provided that it is declared as due in the VAT return. The Polish Article 33a therefore falls within the model provided for by the Directive and does not constitute an exception or an

autonomous derogation from the common EU VAT system [Directive 2006/112/EC].

Article 33a of the VAT Act should be regarded as a special method of accounting for the tax due on the importation of goods, rather than as a tax exemption. The provision allows a VAT-registered taxable person to report import VAT due in the tax return for the period in which the tax liability on the importation arose. In the literature and international practice, this mechanism is referred to as postponed VAT accounting, i.e. accounting for import VAT in the tax return instead of paying it at customs clearance [VAT Act, Article 33a(1); Bartosiewicz 2026].

Access to Article 33a is subject to statutory conditions intended to ensure a minimum level of taxpayer reliability. These include VAT registration, compliance with tax and social-security obligations and, as a rule, the lodging of customs declarations through a direct or indirect representative, subject to exceptions provided for authorised operators and holders of specified customs simplifications [VAT Act, Article 33a; Regulation No 952/2013].

The relationship between Article 33a and the right of deduction is particularly important. Article 86(2)(2)(b) of the VAT Act explicitly provides that, in the case of the importation of goods, input tax also includes the amount of tax due in the case referred to in Article 33a. As a result, for a compliant taxable person using imported goods for taxable activities, the mechanism is generally economically neutral while being highly significant for liquidity [VAT Act, Article 86(1) and Article 86(2)(2)(b)].

Taxable persons importing goods under Article 33a are required to submit monthly tax returns. The legislature therefore compensates for the shift of collection from the border to the tax return by requiring more frequent reporting and, consequently, facilitating subsequent verification [VAT Act, Article 99(3a)(3), (3d) and (4a)].

The mechanism for accounting for the tax due on the importation of goods in the tax return was introduced into the VAT Act on 1 December 2008 by the Act of 7 November 2008, which added Articles 33a-33c after Article 33. Initially, this possibility was linked to specific simplified procedures under customs law [VAT Amendment Act 2008].

A significant liberalisation of Article 33a took effect on 1 July 2020. Pursuant to Article 33(2) of the Act of 31 July 2019 amending certain acts in order to reduce regulatory burdens, the previous link between declaration-based accounting for import VAT and the placing of goods under a specified simplified procedure was replaced by a rule under which the mechanism may be used by a VAT-registered taxable person. At the same time, the legislature introduced a requirement that customs declarations be lodged by a direct or indirect representative, with an exception for taxable persons holding AEO status or authorisations for the customs simplifications referred to in Articles 166 or 182 of the UCC. The possibility of using Article 33a solely in conjunction with a simplified procedure under customs law was therefore abandoned (before the amendment, goods had to be covered in Poland by

a simplification under Article 166 or 182 of the UCC), and the focus shifted to enabling VAT taxpayers to use Article 33a [Act Amending Certain Acts 2019].

METHODOLOGY AND ANALYTICAL ASSUMPTIONS

The study is based on aggregated administrative data concerning imports accounted for under Article 33a of the VAT Act. The data were obtained on request from the Analytical Centre Department of the Tax Administration Chamber in Warsaw (Wydział Centrum Analityczne Izby Administracji Skarbowej w Warszawie), which processes and provides international trade data. The dataset covers the years 2020–2024 and includes import values, the number of entities and customs declarations, commodity codes, mass and geographical information concerning countries of dispatch and origin.

The starting year was selected primarily because a significant liberalisation of Article 33a entered into force on 1 July 2020, substantially broadening access to declaration-based accounting for import VAT. The year 2020 should therefore be regarded as a transition year in the development of the current model. The end of the period, 2024, was the latest complete calendar year available when the dataset was obtained. The five-year period also covers substantial changes in international trade conditions associated with the COVID-19 pandemic, disruptions in global supply chains and the war in Ukraine.

Commodity analysis was conducted at the four-digit Combined Nomenclature level (CN4). The level of aggregation was selected during the data-acquisition process as a compromise between commodity detail and coverage of the entire population of Article 33a imports over the five-year period. Within the scope of the present data-acquisition process, obtaining a complete CN8 dataset for all commodity groups over the five-year period was not practicable; a more detailed extraction would have required prior narrowing of the request to selected commodity groups. Such an approach would reduce the exploratory character of the study by requiring the relevant goods to be selected before the analysis. CN4 therefore makes it possible to analyse the full commodity structure and identify groups that may subsequently justify analysis at CN8/TARIC level.

The analysis combines descriptive statistics, year-on-year changes, comparison with Poland's total imports, commodity concentration analysis and a Pareto-type assessment of the distribution of import value across CN4 headings. For the selected case of CN 2841, geographical concentration was additionally assessed using the Herfindahl-Hirschman Index (HHI). HHI was selected because it provides a simple synthetic measure of concentration and gives greater weight to dominant geographical directions through the squaring of individual shares. In this study, HHI is used as a screening measure rather than as evidence of irregularity.

The study is subject to important data limitations. The dataset is aggregated and does not permit the identification or analysis of individual importers, their compliance histories or links between specific customs declarations and individual

VAT/JPK settlements. Similarly, CN4 data do not distinguish individual products within each four-digit heading. Consequently, changes in import value, unit value or geographical concentration identified in this study should be interpreted as signals that may justify further analysis, not as evidence of fraud or tax abuse. More detailed risk assessment would require access to CN8/TARIC and transaction- or taxpayer-level data, which were not available for the present study.

Since VAT is generally neutral for compliant taxable persons, the main reference point for assessing the economic scale of Article 33a is the total value of Poland's imports rather than State-budget VAT revenue. Any VAT-equivalent calculations presented later in the article are therefore illustrative measures of scale and should not be interpreted as estimates of the VAT gap or potential revenue loss.

SCALE AND DYNAMICS OF IMPORTS UNDER ARTICLE 33A IN 2020-2024

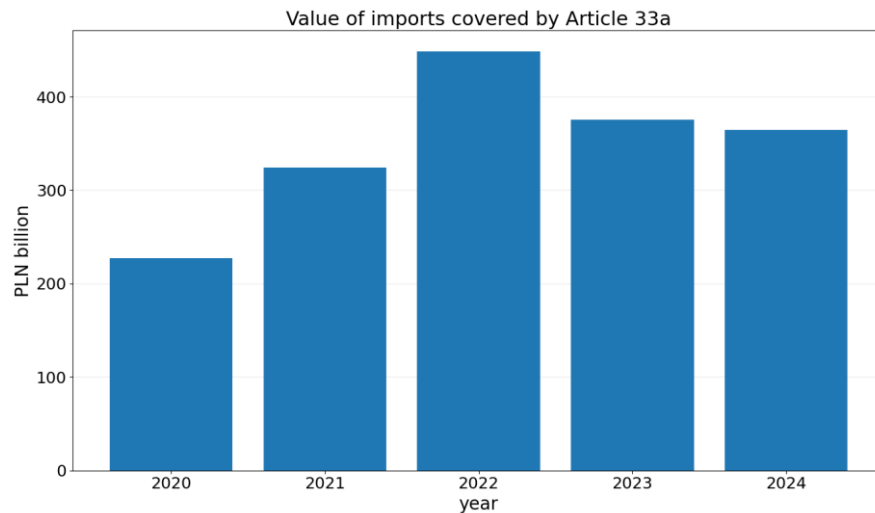
The value of imports covered by the analysed data on the Article 33a mechanism increased from approximately PLN 227.1 billion in 2020 to approximately PLN 448.4 billion in 2022, and subsequently decreased to approximately PLN 364.4 billion in 2024. The decline in value after 2022 does not, however, mean that taxable persons were withdrawing from the mechanism. The number of companies increased from 6,999 to 9,186, while the number of declarations under the simplified import procedure reached its highest level in the series in 2024.

Table 1. Scale and dynamics of imports using Article 33a in 2020-2024

Year	Value (PLN billion)	Value growth y/y	Number of companies	Growth in number of companies y/y	Number of declarations	Growth in number of declarations y/y
2020	227.1	—	6,999	—	905,391	—
2021	323.9	42.6%	7,629	9.0%	1,095,954	21.0%
2022	448.4	38.4%	8,285	8.6%	986,079	-10.0%
2023	375.6	-16.2%	8,702	5.0%	961,639	-2.5%
2024	364.4	-3.0%	9,186	5.6%	1,142,316	18.8%

Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw

Figure 1. Value of imports covered by Article 33a in 2020-2024



Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw

Two main conclusions may be drawn. First, the value of imports under the Article 33a mechanism reached its maximum in 2022, which may be linked both to raw-material prices and supply-chain disruptions, as well as to the fuller use of the simplified import mechanism following its liberalisation in 2020. Second, the steadily increasing number of companies suggests that the mechanism is becoming more widespread, which means that the Article 33a mechanism is not a marginal phenomenon, but a permanent and growing element of import practice.

ARTICLE 33A OF THE VAT ACT AGAINST POLAND'S TOTAL IMPORTS

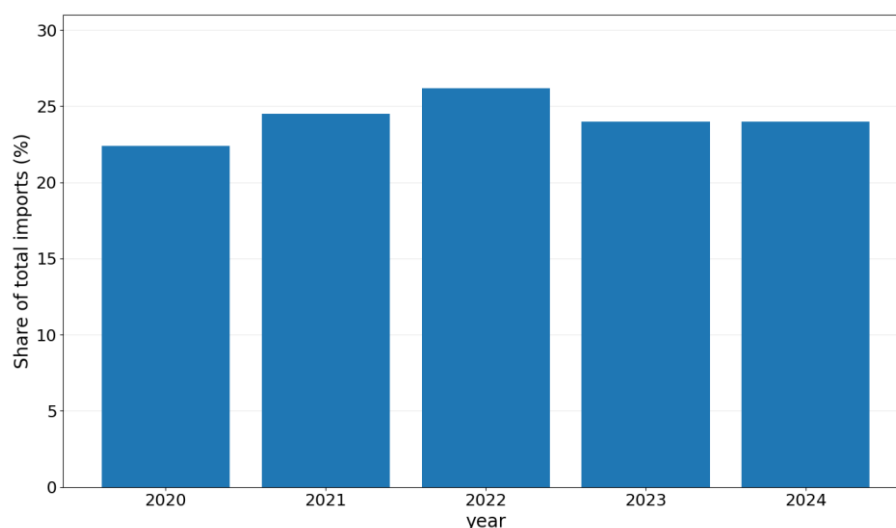
The most appropriate comparison is between the value of imports under the Article 33a mechanism and Poland's total imports. Statistics Poland data were used, according to which Poland's imports at current prices amounted to PLN 1,521.0 billion in 2024, while imports under Article 33a accounted for 24.0% (Table 2). On this basis, Article 33a accounted for approximately 22-26% of the value of Poland's imports during the period under review [GUS 2025; GUS 2020-2023].

Table 2. Article 33a against Poland's total imports in 2020-2024

Year	Poland's total imports (PLN billion)	Imports under Article 33a (PLN billion)	Share of imports under Article 33a in total imports
2020	1,015.4	227.1	22.4%
2021	1,323.0	323.9	24.5%
2022	1,711.8	448.4	26.2%
2023	1,568.3	375.6	24.0%
2024	1,521.0	364.4	24.0%

Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw and Statistics Poland data

Figure 2. Share of imports under Article 33a in the total value of Poland's imports



Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw and Statistics Poland data

The size of the share (Figure 2) illustrates the scale of use of the simplified procedure: the mechanism covers approximately one quarter of the value of Poland's imports. From the perspective of the customs and tax authorities, this means that even though VAT is neutral for a compliant taxable person, the Article 33a mechanism generates a considerable analytical burden. It is necessary to verify not only compliance with the formal conditions, but also the commodity structure, unit values, country of dispatch, country of origin and regularity of the entities' activity—in other words, everything that may raise doubts given the scale of these imports.

COMMODITY STRUCTURE AND CONCENTRATION OF THE VALUE OF IMPORTS UNDER ARTICLE 33A OF THE VAT ACT

The commodity structure of imports under Article 33a was analysed at CN4 level, in accordance with the methodological assumptions described above. The individual CN4 headings were additionally assigned to broader commodity groups corresponding to CN chapters in order to provide a synthetic representation of the import structure.

The dataset contains 1,179 distinct CN4 codes (commodity groups). The top 20% of codes, i.e. 236 items ranked by import value, account for 90.4% of the total value of imports under the simplified import procedure in 2020-2024. This indicates a very strong concentration effect: most of the value generated by the mechanism is attributable to a relatively narrow range of goods.

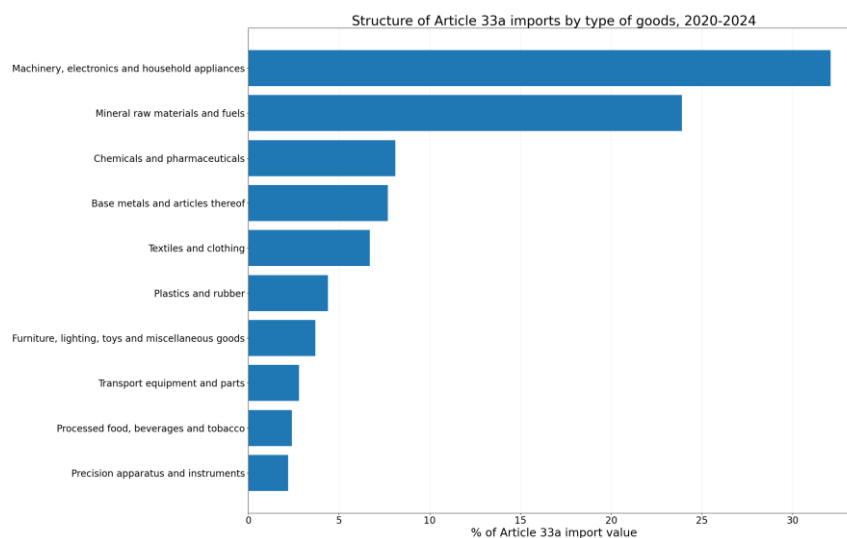
Table 3. Commodity structure and concentration of Article 33a import value by commodity group, 2020-2024

Commodity group	Value 2020-2024 (PLN billion)	Share of imports under Article 33a	Number of CN4 commodity groups	Value in top 20% of CN4	Share of top 20% in total Article 33a imports	CN4 codes in top 20%	Group coverage by top 20%
Machinery, electronics and household appliances	558.6	32.1%	134	540.8	31.1%	61	96.8%
Mineral raw materials and fuels	415.2	23.9%	60	407.7	23.4%	8	98.2%
Chemicals and pharmaceuticals	140.2	8.1%	177	114.2	6.6%	24	81.4%
Base metals and articles thereof	134.3	7.7%	146	110.5	6.4%	33	82.3%
Textiles and clothing	116.7	6.7%	143	94.2	5.4%	27	80.7%
Plastics and rubber	76.6	4.4%	43	70.7	4.1%	20	92.3%

Commodity group	Value 2020-2024 (PLN billion)	Share of imports under Article 33a	Number of CN4 commodity groups	Value in top 20% of CN4	Share of top 20% in total Article 33a imports	CN4 codes in top 20%	Group coverage by top 20%
Furniture, lighting, toys and miscellaneous goods	63.5	3.7%	35	58.9	3.4%	9	92.8%
Transport equipment and parts	47.9	2.8%	39	40.6	2.3%	7	84.9%
Processed food, beverages and tobacco	42.1	2.4%	52	35.9	2.1%	8	85.3%
Precision apparatus and instruments	38.1	2.2%	55	30.4	1.8%	9	80.0%

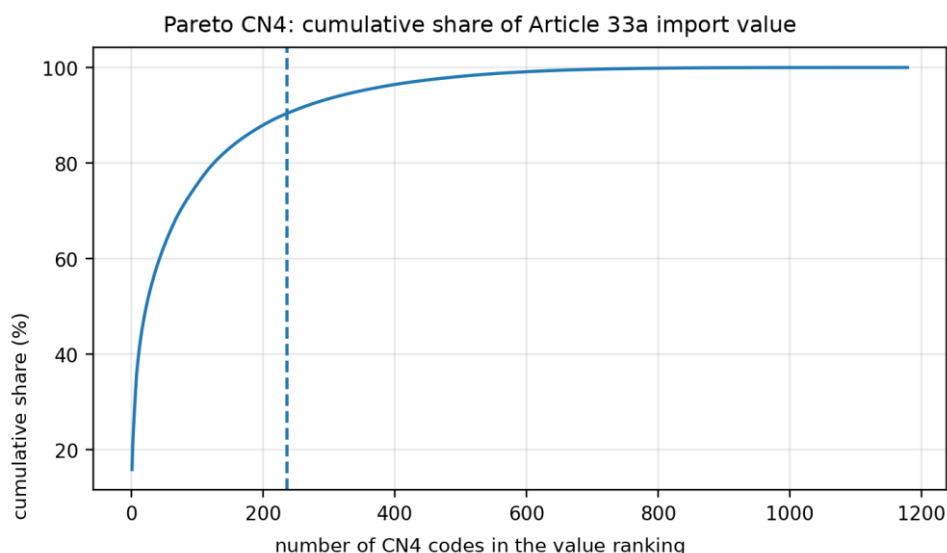
Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw

Figure 3. Structure of import value under Article 33a by type of goods, 2020-2024



Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw

Figure 4. Pareto distribution of CN4 codes: cumulative share of Article 33a import value



Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw

The result is fairly unambiguous. The largest commodity group imported into Poland under the Article 33a mechanism is not food, but machinery, electronics and household appliances, which account for approximately 32.1% of the value of imports under Article 33a. The second group is mineral raw materials and fuels (23.9%). Processed food, beverages and tobacco are of much less importance (approximately 2.4%), while vegetable and animal products together do not exceed 1% of the value. In practice, Article 33a is therefore particularly important for importers of electronics, machinery, fuels, chemicals, metals, textiles and industrial parts.

This is also confirmed by individual CN4 codes. CN 2709, i.e. crude petroleum oils, represented the highest import value during the period and accounted for 15.9% of all simplified imports. It was followed at a considerable distance by CN 8471, i.e. automatic data-processing machines, with 5% of the total, and CN 2710, i.e. petroleum products, with 2.9%. This demonstrates the clearly industrial, raw-material-based and technological structure of imports under the Article 33a mechanism.

Table 4. CN4 headings with the highest value of imports covered by Article 33a, 2020-2024

CN4	Short description	Group	Value (PLN billion)	Share	Cumulative share
2709	Petroleum oils and oils obtained from bituminous minerals, crude	Mineral raw materials and fuels	276.3	15.9%	15.9%
8471	Automatic data-processing machines and units thereof; magnetic or optical readers; machines for transcribing...	Machinery, electronics and household appliances	86.8	5.0%	20.9%
2710	Petroleum oils and oils obtained from bituminous minerals, other than crude; preparations...	Mineral raw materials and fuels	49.8	2.9%	23.7%
2701	Coal; briquettes, ovoids and similar solid fuels manufactured from coal	Mineral raw materials and fuels	48.0	2.8%	26.5%
8517	Telephone sets and other apparatus for the transmission or reception of voice, images or other data...	Machinery, electronics and household appliances	44.1	2.5%	29.0%
8529	Parts suitable for use solely or principally with the apparatus of headings 8524 to 8528	Machinery, electronics and household appliances	41.6	2.4%	31.4%
8507	Electric accumulators, including separators therefor, whether or not rectangular	Machinery, electronics and household appliances	39.6	2.3%	33.7%
2841	Salts of oxometallic or peroxometallic acids	Chemicals and pharmaceuticals	38.2	2.2%	35.9%

CN4	Short description	Group	Value (PLN billion)	Share	Cumulative share
8708	Parts and accessories of the motor vehicles of headings 8701 to 8705	Transport equipment and parts	24.7	1.4%	37.3%
8528	Monitors and projectors; reception apparatus for television...	Machinery, electronics and household appliances	23.1	1.3%	38.7%
8524	Flat panel display modules, whether or not incorporating touch-sensitive screens	Machinery, electronics and household appliances	22.4	1.3%	39.9%
2304	Oilcake and other solid residues, whether or not ground or in pellets, resulting from the extraction...	Processed food, beverages and tobacco	20.6	1.2%	41.1%
8544	Insulated wire, cable and other insulated electric conductors; optical-fibre cables...	Machinery, electronics and household appliances	19.3	1.1%	42.2%
8411	Turbo-jets, turbo-propellers and other gas turbines	Machinery, electronics and household appliances	16.9	1.0%	43.2%
8504	Electrical transformers, static converters (e.g. rectifiers) and inductors	Machinery, electronics and household appliances	16.8	1.0%	44.2%

Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw

SELECTED CASE: CN 2841 - SALTS OF OXOMETALLIC OR PEROXOMETALLIC ACIDS

CN code 2841 was selected from the broader CN4 dataset as an illustrative case for more detailed analysis because it combines three characteristics relevant to risk profiling: a sudden increase in import value, a high unit value and strong

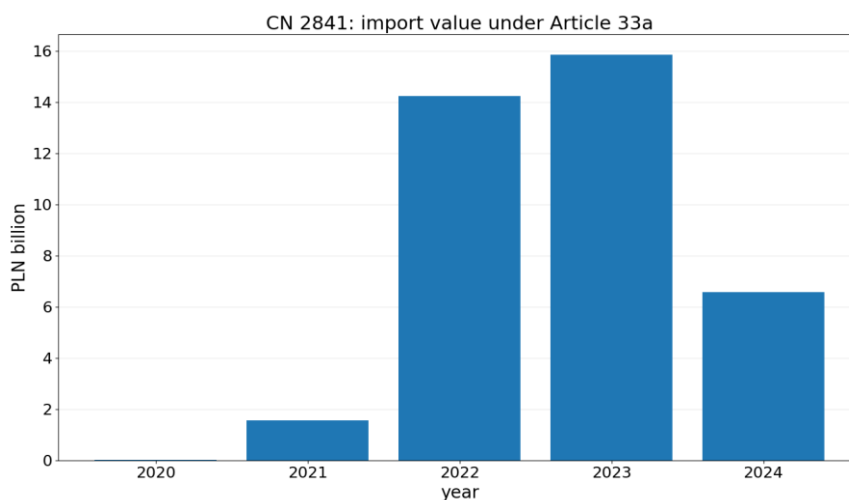
geographical concentration. In 2020, imports under this heading using Article 33a were marginal, amounting to approximately PLN 16 million, but in 2022–2023 they reached several billion zlotys per year, including almost PLN 16 billion in 2023.

Table 5. CN 2841 under Article 33a - value, mass, unit value and illustrative VAT equivalent

Year	CN 2841 value (PLN billion)	Mass (thousand tonnes)	Unit value (PLN/kg)	Share in the Article 33a procedure	Share in Poland's imports	Illustrative VAT equivalent at 23% (PLN billion)	CN 2841 value growth y/y
2020	0.02	1.87	8.62	0.01%	0.00%	0.00	—
2021	1.56	18.41	84.82	0.48%	0.12%	0.36	9,590.8%
2022	14.24	74.56	191.04	3.18%	0.83%	3.28	812.3%
2023	15.85	76.84	206.30	4.22%	1.01%	3.65	11.3%
2024	6.57	55.96	117.36	1.80%	0.43%	1.51	-58.6%

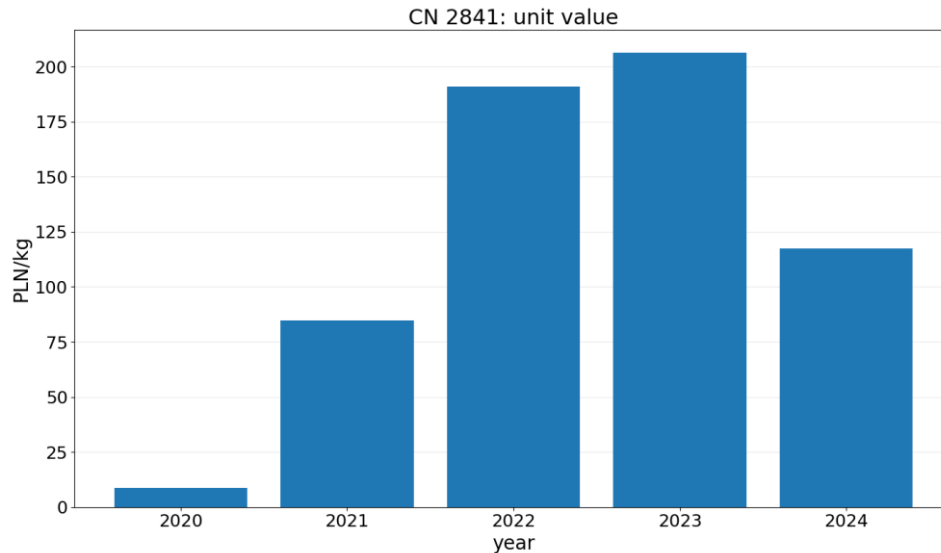
Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw

Figure 5. CN 2841: import value under Article 33a (PLN billion), 2020-2024



Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw

Figure 6. CN 2841: unit value (PLN/kg), 2020-2024



Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw

The most likely hypothesis concerning the sharp increase in the value of imports under CN code 2841 is linked to the development of the lithium-ion battery supply chain in Poland and Europe. CN heading 2841 includes, among other products, salts of oxometallic and peroxometallic acids, while at CN/TARIC level 28419040 lithium nickel manganese cobalt oxides are distinguished, including, for example, NMC/NCM cathode powders used in lithium-ion batteries [TariffNumber 2026].

This conclusion is consistent with the development of the battery sector in Poland. The LG Energy Solution Wrocław plant began production in 2017 and, according to a Reuters overview, its target production capacity was expected to reach 115 GWh by 2025, which may indicate an increase in production during the period under discussion [Reuters 2024]. At the same time, the European Investment Bank announced in 2020 that it was financing the expansion of LG Chem's battery production in Poland, and in 2022 the European Commission approved Polish State aid for the expansion of the LG Chem/LG Energy Solution electric-vehicle battery plant near Wrocław [European Investment Bank 2020; European Commission 2022].

During the same period, Umicore was developing the European cathode-material supply chain in Nysa and subsequently, together with PowerCo/Volkswagen, announced the Ionway project concerning the production of cathode materials in Poland [Umicore 2023; Ionway 2023].

The increase in the value of imports of goods under CN 2841 may therefore reflect a genuine economic change and, consequently, imports of cathode materials or specialised metal compounds required for battery production. This does not therefore indicate abuse; aggregated CN4-level data instead point to a profile requiring in-depth analysis. They thus provide a useful starting point for developing a post-clearance control profile.

HHI AS A SIMPLE MEASURE OF THE CONCENTRATION OF CN 2841 IMPORT VALUE

The Herfindahl-Hirschman Index (HHI) was used to assess the geographical concentration of imports of goods under CN 2841. In its conventional application, the HHI measures market concentration and is calculated as the sum of the squared shares of individual market participants [Czarny, Rapacki 2002]. In this study, the same method was applied not to company shares, but to the shares of countries of dispatch or countries of origin in the import value of CN 2841. The index was calculated using the following formula:

$$HHI = \sum_{i=1}^n s_i^2,$$

where: s_i - the share of the i -th country in the import value of the commodity heading under examination.

If the shares are expressed as decimal fractions, the HHI takes values ranging from close to 0 to 1. A value close to 1 indicates extreme concentration, i.e. the dominance of a single geographical direction, while lower values indicate greater dispersion of imports among many countries. Alternatively, shares may be expressed in percentage points, in which case the HHI ranges from 0 to 10,000 [Dynowska, Kes 2016]. This study uses the fractional variant, i.e. a range from 0 to 1.

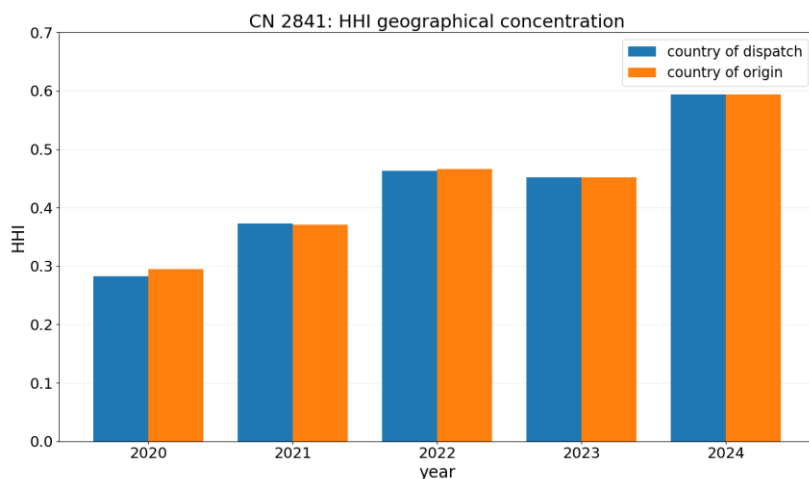
High concentration may result from lawful and economically rational factors, such as supplier specialisation, the location of cathode-material producers or the organisation of the supply chain. The index should therefore be treated not as evidence of abuse, but as a tool for selecting areas requiring in-depth analysis. Its use in the case examined is additionally justified because CN 2841 is characterised by a high unit value and a sharp increase in import value in 2021-2023. Where such an increase is also strongly geographically concentrated, the customs and tax administration should consider moving to the operational-data level of CN8/TARIC codes, customs value, transport documents, links between the parties and the subsequent use of the goods in production in order to conduct more precise risk profiling for post-clearance import controls.

Table 6. Geographical concentration of imports of CN 2841 goods according to the HHI

Year	HHI - country of dispatch	Top 3 dispatch countries	HHI - country of origin	Top 3 origin countries
2020	0.283	83.2%	0.295	83.0%
2021	0.373	99.3%	0.371	99.3%
2022	0.463	99.9%	0.466	99.9%
2023	0.452	99.9%	0.452	99.9%
2024	0.594	99.9%	0.594	99.9%

Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw

Figure 7. CN 2841: geographical concentration according to the HHI



Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw

Table 7. Main countries of dispatch - CN 2841 in 2020-2024

Country of dispatch for CN 2841	Import value 2020-2024 (PLN billion)	Share
Republic of Korea	24.32	63.6%
Japan	7.28	19.0%
China	6.58	17.2%
India	0.03	0.1%
South Africa	0.01	0.0%

Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw

The HHI values show that after 2020 CN 2841 was very strongly concentrated geographically. In 2021-2024, the share of the three main directions generally exceeded 99%. The data therefore do not concern small, dispersed trade, but a specialised flow linked to a specific supply chain.

ILLUSTRATIVE VAT EQUIVALENT UNDER THE GENERAL IMPORT-ACCOUNTING RULES AND ITS RELATIVE SCALE

Only after presenting the value of imports under Article 33a against Poland's total imports can the analysed mechanism be cautiously related to the VAT dimension. Table 8 does not present the actual amount of "import VAT" or actual budget revenue associated with the application of Article 33a. In order to illustrate the financial scale of the analysed flows, an illustrative VAT equivalent under the general import-accounting rules was nevertheless calculated by applying, for illustrative purposes, a rate of 23% to the value of imports covered by Article 33a of the VAT Act.

The value calculated in this manner should not, however, be equated with the amount of tax that would actually have been paid into the budget if the Article 33a mechanism had not been applied, because the actual taxable amount for VAT on importation need not be identical to the import value used in this estimate. Furthermore, individual goods may be subject to different tax rates. Another important aspect must also be taken into account: abandoning accounting under Article 33a could affect the direction of trade flows, including by causing import customs clearance to be carried out in other EU Member States.

Table 8. Illustrative VAT equivalent under the general import-accounting rules and its relative-scale indicator

Year	Imports under Article 33a (PLN billion)	Illustrative VAT equivalent under the general import-accounting rules - 23% (PLN billion)	State-budget VAT revenue (PLN billion)	Relative-scale indicator for the illustrative VAT equivalent
2020	227.1	52.2	184.5	28.3%
2021	323.9	74.5	215.7	34.5%
2022	448.4	103.1	230.5	44.7%
2023	375.6	86.4	244.3	35.4%
2024	364.4	83.8	287.7	29.1%

Source: own calculations based on data from the Analytical Centre Department of the Tax Administration Chamber in Warsaw

The comparison of the estimated amount with total State-budget VAT revenue is therefore purely illustrative. On this basis, a relative-scale indicator for the illustrative VAT equivalent was calculated, showing the relationship between the

illustrative VAT equivalent corresponding to the value of imports covered by the simplified import procedure and annual State-budget VAT revenue. This indicator is not a measure of the tax gap, potential revenue loss or the actual risk of revenue loss; it serves solely to illustrate the scale of the tax value associated with flows covered by the Article 33a mechanism against the background of the entire VAT revenue system [Ministry of Finance 2020-2024].

In 2022, the illustrative VAT equivalent calculated by applying a 23% rate to the value of imports accounted for under Article 33a amounted to approximately PLN 103.1 billion, equivalent to 44.7% of State-budget VAT revenue in that year. As explained above, this ratio does not represent VAT generated by Article 33a, potential budget revenue or a VAT-gap estimate. It serves only to illustrate the financial scale of import flows subject to declaration-based accounting and, consequently, the potential importance of effective ex post risk profiling.

DISCUSSION AND CONCLUSIONS

The results show that Article 33a was a significant component of Polish import practice in 2020–2024 following the liberalisation of the mechanism in 2020. Its importance should be assessed not only from the perspective of liquidity facilitation for businesses but also from the perspective of customs and tax administration. A mechanism covering approximately one quarter of the value of Polish imports transfers a substantial volume of transactions from immediate collection of import VAT towards declaration-based accounting and subsequent verification.

The findings support the broader literature emphasising the increasing importance of data-based approaches to tax and customs risk management. Alexopoulos et al. [2025] demonstrate the usefulness of transaction and network data in VAT-risk detection, while Park, Kim and Cha [2022] and Mai et al. [2021] show the relevance of changing trade patterns for customs-risk analysis. The present study complements this literature by showing how relatively simple measures applied to aggregated Article 33a data—changes in import value, commodity concentration, unit values and geographical concentration—may serve as an initial screening layer before more detailed operational analysis.

The CN 2841 case illustrates both the usefulness and the limits of this approach. A rapid increase in value combined with high unit values and strong geographical concentration may justify further examination, but it does not constitute evidence of abuse. In this case, developments in the European lithium-ion battery supply chain provide an economically plausible explanation for the observed trade pattern. Risk profiling should therefore be understood as a mechanism for selecting areas for further verification rather than as a method for automatically classifying unusual trade flows as fraudulent.

From the perspective of tax administration, the results suggest that monitoring Article 33a should combine changes in commodity flows with information on unit

values, countries of dispatch and origin and geographical concentration. Where significant deviations are identified, analysis may subsequently move from aggregated CN4 data to CN8/TARIC codes and, where legally and technically available, to transaction-level customs data, VAT/JPK information and taxpayer compliance history. Such a multi-stage approach may also reduce the risk that static profiles become ineffective when trade patterns or commodity classifications change over time.

The study has several limitations. The data are aggregated and therefore do not allow individual importers to be identified, nor do they permit direct links to be established between customs declarations and the VAT settlements of individual taxpayers. The CN4 level also does not identify specific products within individual headings. For these reasons, the study cannot estimate the scale of actual abuse or the VAT gap associated with Article 33a.

Future research could extend the analysis using CN8/TARIC and transaction-level data for commodity groups identified through initial screening, as well as by linking customs and tax data where access conditions permit. Comparative analysis of postponed VAT accounting in other EU Member States would also help assess whether the patterns observed in Poland are specific to Article 33a or reflect broader characteristics of declaration-based import VAT accounting.

REFERENCES

- Act of 11 March 2004 on Tax on Goods and Services (consolidated text: Journal of Laws of 2025, item 775, as amended) (in Polish).
- Act of 31 July 2019 Amending Certain Acts in Order to Reduce Regulatory Burdens (in Polish).
- Act of 7 November 2008 Amending the Act on Tax on Goods and Services and Certain other Acts (in Polish).
- Alexopoulos A., Dellaportas P., Gyoshev S., Kotsogiannis C., Olhede S.C., Pavkov T. (2025) A Network Approach to Detect Value Added Tax fraud. *Journal of the Royal Statistical Society Series A: Statistics in Society*, qnaf205, <https://doi.org/10.1093/jrssa/qnaf205>
- Analytical Centre Department of the Tax Administration Chamber in Warsaw, administrative data supplied to the author on request.
- Bartosiewicz A. (2026) VAT. Komentarz 2026, Vol. 1, Article 33a, Wolters Kluwer Polska (in Polish).
- Council Directive 2006/112/EC of 28 November 2006 on the Common System of Value Added Tax.
- Czarny B., Rapacki R. (2002) *Podstawy ekonomii*, 2nd revised ed., Polskie Wydawnictwo Ekonomiczne, Warszawa (in Polish).
- Dynowska J., Kes Z. (2016) Pomiar zróżnicowania odchyleń w kontroli budżetowej. *Finanse, Rynki Finansowe, Ubezpieczenia*, 2(80), Part 2, 325-335, 10.18276/frfu.2016.2.80/2-34 (in Polish).

- European Commission (2022) State Aid: Commission Approves €95 Million Polish Aid for Expansion of LG Chem's Electric-Vehicle Battery Plant, https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1861
- European Investment Bank (2020) EBI wspiera produkcję baterii do pojazdów elektrycznych w Europie, przyznając polskiemu przedsiębiorstwu LG Chem Wrocław Energy kredyt w wysokości 480 mln EUR, 26.03.2020, <https://www.eib.org/en/press/all/2020-088-electric-vehicle-battery-production-in-europe-gets-boost-thanks-to-eib-loan-of-eur480-million-to-lg-chem-wroclaw-energy-in-poland?lang=pl> (in Polish).
- GUS (2020-2023) Archive of Statistical Releases on Foreign Trade in Goods for 2020-2023 (in Polish).
- GUS (2025) Obroty towarowe handlu zagranicznego ogółem i według krajów (dane ostateczne) w 2024 r, <https://stat.gov.pl/obszary-tematyczne/ceny-handel/handel/obroty-towarowe-handlu-zagranicznego-ogolem-i-wedlug-krajow-dane-ostateczne-w-2024-r-2,17.html> (in Polish).
- HMRC (2025), Guidance: Check When You Can Account for Import VAT on Your VAT Return, <https://www.gov.uk/guidance/check-when-you-can-account-for-import-vat-on-your-vat-return>
- Ionway (2023) IONWAY Locates Its First CAM Production Plant in Nysa, Poland, 7.10.2023, <https://ionway.com/blog/2023/10/07/ionway-to-locate-its-first-production-plant-in-nysa-poland/>
- Keen M., Smith S. (2007) VAT Fraud and Evasion: What Do We Know and What Can Be Done? IMF Working Paper WP/07/31, <https://www.imf.org/external/pubs/ft/wp/2007/wp0731.pdf>
- Krychowski A. (2023) Uproszczenie przy imporcie towarów wynikające z art. 33a ustawy o podatku od towarów i usług. Wrocławsko-Lwowskie Zeszyty Prawnicze, 14, <https://doi.org/10.19195/2082-4939.14.14> (in Polish).
- Laszuk M., Komorowski E., Michalski R. (2022) Prawo celne. Komentarz, Wolters Kluwer Polska, Warszawa (in Polish).
- Mai T.-D., Hoang K., Baigutanova A., Alina G., Kim S. (2021) Customs Fraud Detection in the Presence of Concept Drift, 370-379, 10.1109/ICDMW53433.2021.00052
- Markowicz I., Baran P. (2021) Mirror Data Asymmetry in International Trade by Commodity Group: The Case of Intra-Community Trade. *Oeconomia Copernicana*, 12(4), 889-905, 10.24136/oc.2021.029
- Mastalski R. (2021) Prawo podatkowe, 12th ed., C.H. Beck, Warszawa (in Polish).
- Matarewicz J. (2023) Ustawa o podatku od towarów i usług. Komentarz aktualizowany, LEX/el, <https://sip.lex.pl/komentarze-i-publicacje/komentarze/ustawa-o-podatku-od-towarow-i-uslug-komentarz-aktualizowany-587754168> (in Polish).
- Michalik T. (2024) VAT 2024. Komentarz. C.H. Beck (in Polish).
- Ministry of Finance (2020-2024) Szacunkowe wykonanie budżetu państwa, annual estimates of State-budget execution for 2020-2024, <https://www.gov.pl/web/finanse/szacunkowe-wykonanie-budzetu> (in Polish).
- OECD (2017) International VAT/GST Guidelines Report, https://www.oecd.org/en/publications/international-vat-gst-guidelines_9789264271401-en.html

- Park S., Kim S., Cha M. (2022) Knowledge Sharing via Domain Adaptation in Customs Fraud Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(11), 12062–12070, <https://doi.org/10.1609/aaai.v36i11.21465>
- Regulation (EU) No 952/2013 of the European Parliament and of the Council of 9 October 2013 Laying down the Union Customs Code; in particular Articles 38, 166 and 182 were used, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32013R0952>
- Reuters (2024) Companies Investing in EV Battery Factories in Europe, 10.12.2024, <https://www.reuters.com/business/autos-transportation/companies-invest-ev-battery-factories-europe-2023-05-18/>
- Smoleń P., Wójtowicz W. (Eds.) (2021) *Prawo podatkowe*, 5th ed., Wydawnictwo C.H. Beck, Warszawa (in Polish).
- TariffNumber (2026) HS Code 28419040 - Lithium Nickel Manganese Cobalt Oxides, <https://www.tariffnumber.com/2026/28419040>
- Umicore (2023) Umicore Accelerates European E-Mobility with Nysa Gigafactory, 21.03.2023, <https://www.umicore.com/en/newsroom/news/umicore-accelerates-european-e-mobility-with-nysa-gigafactory/>

ALFABETYZM STATYSTYCZNY A PUŁAPKI MYŚLENIA. ANALIZA LUKI KOMPETENCYJNEJ NA STYKU EDUKACJI ŚREDNIEJ I WYŻSZEJ

Ewa Wasilewska  <https://orcid.org/0000-0001-9065-5526>

Instytut Ekonomii i Finansów

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

e-mail: ewa_wasilewska@sggw.edu.pl

Streszczenie: Celem artykułu jest ocena poziomu alfabetyzmu statystycznego uczniów szkół średnich i studentów oraz analiza wybranych komponentów tej kompetencji. Badanie przeprowadzono wśród 432 uczniów i studentów z wykorzystaniem autorskiego testu sytuacyjnego. Wyniki wskazują na istnienie luki kompetencyjnej pomiędzy efektami formalnej edukacji matematyczno-statystycznej a poziomem alfabetyzmu statystycznego ujawnianym podczas interpretacji danych i rozumowania probabilistycznego. Analiza ujawniła anomalię edukacyjną: uczniowie realizujący rozszerzony program matematyki wykazali wyższy poziom alfabetyzmu statystycznego niż studenci po akademickim kursie statystyki.

Słowa kluczowe: alfabetyzm statystyczny, luka kompetencyjna, edukacja statystyczna, heurystyki poznawcze, rozumowanie probabilistyczne, interpretacja danych

JEL classification: A21, A22, D91, C12, I21

WPROWADZENIE

We współczesnej gospodarce opartej na wiedzy i danych zdolność do rozumienia, interpretowania oraz krytycznej oceny informacji liczbowych staje się jedną z podstawowych kompetencji obywatelskich. W odpowiedzi na te wyzwania coraz większego znaczenia nabiera pojęcie alfabetyzmu statystycznego (statistical literacy), rozumianego jako zdolność do odczytywania, interpretowania, krytycznej oceny oraz komunikowania informacji statystycznych wykorzystywanych w życiu codziennym [Gal 2002]. W literaturze przedmiotu pojęcie to bywa zestawiane z pokrewnymi, lecz odrębnymi konstruktami – rozumowaniem statystycznym

<https://doi.org/10.22630/MIBE.2026.27.3.12>



(statistical reasoning) oraz myśleniem statystycznym (statistical thinking), przy czym nie ma pełnej zgodności co do precyzyjnych granic pomiędzy tymi pojęciami [Ben-Zvi i Garfield 2004]. Wysoki poziom alfabetyzmu statystycznego społeczeństwa stanowi istotny warunek podejmowania racjonalnych decyzji mikro- i makroekonomicznych. Jak podkreśla Oleński [2020], jakość statystyki publicznej oraz środowiska informacyjnego, w którym ona funkcjonuje, pełni funkcję istotnej bariery chroniącej przed dezinformacją i manipulowaniem danymi statystycznymi w przestrzeni publicznej. Rosnąca rola danych we współczesnym społeczeństwie, napędzana rozwojem technologii Big Data, dodatkowo wzmacnia potrzebę rozwijania kompetencji niezbędnych do krytycznej oceny i poprawnej interpretacji wyników badań [Szreder 2016].

Pomimo dynamicznego rozwoju narzędzi analitycznych, rzeczywista zdolność społeczeństwa do poprawnego i refleksyjnego rozumowania probabilistycznego wciąż pozostaje istotnym wyzwaniem. Przełomowe badania Amosa Tversky'ego i Daniela Kahnemana [1974], które dały początek współczesnej ekonomii behawioralnej, wykazały, że ludzie w warunkach niepewności często posługują się heurystykami poznawczymi – uproszczonymi regułami wnioskowania, które choć wymagają minimalnego nakładu wysiłku umysłowego i ułatwiają podejmowanie decyzji, mogą prowadzić do systematycznych błędów i złudzeń probabilistycznych. W realiach ekonomicznych brak świadomej kontroli nad tymi automatyzmami prowadzi do opisywanych w literaturze błędów poznawczych [Tversky i Kahneman 1974], których występowanie może utrudniać właściwą interpretację danych zarówno w życiu codziennym, jak i w praktyce gospodarczej.

Szczególnie krytycznym etapem na ścieżce edukacyjnej jest styk edukacji średniej i wyższej. Tradycyjny model dydaktyki nauk ilościowych kładzie silny nacisk na formalizm matematyczny, algorytmizację oraz pamięciowe opanowanie procedur obliczeniowych. Można jednak przypuszczać, że techniczna zdolność sprawnego posługiwania się aparatem matematycznym nie współwystępuje automatycznie z intuicją statystyczną i refleksyjną odpornością na zniekształcenia poznawcze. Zasadne staje się zatem zbadanie, w jakim stopniu formalna edukacja matematyczno-statystyczna sprzyja rozwojowi alfabetyzmu statystycznego oraz ogranicza podatność na błędy poznawcze ujawniające się podczas rozwiązywania problemów wymagających interpretacji danych i oceny niepewności.

PRZEGLĄD LITERATURY

Alfabetyzm statystyczny w procesie edukacji

Przywołana we Wprowadzeniu koncepcja Gala [2002] nosi pełną nazwę alfabetyzmu statystycznego dorosłych (Adult Statistical Literacy) i zakłada współwystępowanie dwóch komplementarnych komponentów. Pierwszy, poznawczy, obejmuje wiedzę statystyczną, umiejętności matematyczne oraz zdolność krytycznego myślenia. Drugi, dyspozycyjny, obejmuje postawy,

przekonania i motywację do podejmowania refleksyjnej oceny informacji. Rolę alfabetyzmu statystycznego w rozwoju kultury statystycznej oraz świadomym funkcjonowaniu społeczeństwa podkreślają również polscy autorzy, m.in. Domański [2016] i Oleński [2020].

Wraz z dynamicznym rozwojem technologii cyfrowych oraz rosnącą dostępnością dużych zbiorów danych, tradycyjne rozumienie alfabetyzmu statystycznego bywa obecnie rozszerzane o kompetencje określane jako data literacy, których rozwój Gould [2017] postuluje uwzględniać już na wczesnych etapach edukacji, między innymi poprzez wprowadzanie programów data science w szkołach średnich. Na gruncie polskim podobne stanowisko prezentują Jasiewicz i Stopnicka [2016], wskazując, że umiejętność krytycznej analizy danych staje się jednym z podstawowych elementów współczesnych kompetencji obywatelskich i zawodowych.

Jednocześnie liczne badania z zakresu edukacji statystycznej podkreślają, że opanowanie procedur obliczeniowych nie jest tożsame z rozwojem kompetencji związanych z interpretacją danych i rozumowaniem statystycznym. Batanero i współautorzy [2016] wskazują, że skuteczne kształcenie statystyczne wymaga rozwijania nie tylko wiedzy proceduralnej, lecz również rozumienia pojęć oraz umiejętności interpretacji wyników analiz. Podobne stanowisko prezentują Wild i Pfannkuch [1999], którzy podkreślają znaczenie kontekstu oraz procesu wnioskowania statystycznego. Ujęcie to znajduje również uzasadnienie w psychologii poznawczej, gdzie teoria dwuprocesowego przetwarzania informacji zakłada istnienie dwóch komplementarnych sposobów rozumowania: szybkiego, intuicyjnego i automatycznego (System 1) oraz wolniejszego, analitycznego i świadomie kontrolowanego (System 2). Dominacja przetwarzania charakterystycznego dla Systemu 1 sprzyja wykorzystywaniu heurystyk, które mogą prowadzić do błędów poznawczych podczas oceny danych i rozumowania probabilistycznego.

Nowsze badania empiryczne potwierdzają, że rozwój alfabetyzmu statystycznego na kolejnych etapach kształcenia nie ma charakteru jednorodnego. Kurnia, Lowrie i Patahuddin [2024] wykazali różnice między uczniami z różnych poziomów edukacyjnych, wskazując zarazem, że znaczna część badanych posługiwała się wiedzą statystyczną w sposób niekrytyczny. Z kolei Phadke, Beckman i Lock Morgan [2024] wykazali, że kontekst zadań może wpływać na wyniki pomiaru alfabetyzmu statystycznego studentów, co podkreśla znaczenie osadzania diagnozy tej kompetencji w realistycznych sytuacjach problemowych.

Przytoczone koncepcje wskazują, że alfabetyzm statystyczny wykracza poza sprawność wykonywania obliczeń i obejmuje zdolność poprawnego rozumowania w sytuacjach wymagających interpretacji danych oraz oceny niepewności. W praktyce oznacza to konieczność uwzględniania podczas kształcenia nie tylko aspektów rachunkowych, lecz również umiejętności krytycznej interpretacji i analizy informacji statystycznych.

Pułapki myślenia w rozumowaniu probabilistycznym – rola heurystyk poznawczych

Konsekwencją dominacji Systemu 1, opisanej w poprzednim podrozdziale, są systematyczne błędy poznawcze, które w niniejszym artykule określono jako pułapki myślenia [Kahneman 2011]. W literaturze przedmiotu szczególnie często opisywane są cztery rodzaje błędów poznawczych mających znaczenie dla rozumowania statystycznego. Jednym z nich jest niewrażliwość na wielkość próby, polegająca na traktowaniu niewielkich prób jako reprezentatywnych dla całej populacji, mimo że są one znacznie bardziej podatne na przypadkowe odchylenia od wartości oczekiwanych [Tversky i Kahneman 1971]. Istotnym zniekształceniem poznawczym jest również złudzenie gracza, przejawiające się błędnym przekonaniem, że losowe sekwencje zdarzeń powinny w krótkim okresie samoczynnie wyrównywać wcześniejsze odchylenia, co prowadzi do naruszenia zasady niezależności zdarzeń [Kahneman i Tversky 1972]. Do często analizowanych zniekształceń należy także niewrażliwość na zmienność danych, polegająca na koncentrowaniu uwagi wyłącznie na miarach położenia, z pominięciem informacji o rozproszeniu danych, co może prowadzić do błędnej oceny ryzyka i stabilności analizowanych zjawisk [Ben-Zvi i Garfield 2004]. Szczególne miejsce w literaturze zajmuje również błąd proporcji podstawowej (base rate fallacy), polegający na pomijaniu informacji o częstości występowania zjawiska na rzecz pojedynczych informacji szczegółowych, mimo że to właśnie proporcja podstawowa stanowi kluczowy element poprawnego rozumowania probabilistycznego [Bar-Hillel 1980]. Badania Gigerenzer'a i Hoffrage'a [1995] pokazują, że błąd ten występuje nawet u osób posiadających wysokie kwalifikacje zawodowe.

Opisane błędy poznawcze, będące konsekwencją działania heurystyk, stanowiły teoretyczną podstawę konstrukcji dystraktorów wykorzystanych w autorskim narzędziu badawczym. Celem badania nie był pomiar samych heurystyk, lecz ocena wybranych komponentów alfabetyzmu statystycznego w sytuacjach, w których heurystyki mogły prowadzić do wyboru niepoprawnych odpowiedzi.

CEL I METODYKA BADAŃ

Cel badań

Celem badania jest ocena poziomu alfabetyzmu statystycznego oraz identyfikacja ewentualnej luki kompetencyjnej między nim a poziomem formalnej edukacji matematyczno-statystycznej. Na potrzeby badania lukę tę zoperacjonalizowano jako brak oczekiwanego, jednokierunkowego wzrostu wyników testu wraz z kolejnymi poziomami formalnej edukacji matematyczno-statystycznej, ze szczególnym uwzględnieniem porównania studentów po ukończeniu akademickiego kursu statystyki z uczniami realizującymi matematykę w zakresie podstawowym i rozszerzonym. W ramach badania analizowano również

podatność respondentów na wybrane błędy poznawcze ujawniające się podczas rozumowania probabilistycznego. Szczególną uwagę poświęcono umiejętności poprawnej interpretacji informacji statystycznych w sytuacjach wymagających oceny ryzyka, niepewności oraz prawdopodobieństwa.

W nawiązaniu do omówionych wcześniej pułapek myślenia wynikających z działania heurystyk poznawczych analizowano w szczególności, czy wzrost poziomu formalnego wykształcenia matematyczno-statystycznego wiąże się z ograniczeniem podatności na wybrane błędy poznawcze opisywane w literaturze dotyczącej rozumowania probabilistycznego.

W związku z tak sformułowanym celem postawiono następujące pytania badawcze:

- PB1. Czy poziom alfabetyzmu statystycznego różni się pomiędzy uczniami szkół ponadpodstawowych realizującymi matematykę na poziomie podstawowym i rozszerzonym oraz studentami kierunków ekonomiczno-menedżerskich?
- PB2. Które obszary alfabetyzmu statystycznego sprawiają największe trudności poszczególnym grupom respondentów?
- PB3. Czy wraz ze wzrostem poziomu formalnej edukacji matematyczno-statystycznej zmniejsza się podatność na wybrane błędy rozumowania probabilistycznego?

Materiał badawczy i przebieg badania

Badanie zrealizowano w okresie maj–czerwiec 2026 roku metodą sondażu diagnostycznego z wykorzystaniem celowego doboru próby ($N = 432$ respondentów). Strukturę próby tworzą trzy grupy:

- Uczniowie (matematyka w zakresie podstawowym), $N = 123$: uczniowie klas trzecich (przedmaturalnych) dwóch publicznych liceów ogólnokształcących, realizujący matematykę w zakresie podstawowym.
- Uczniowie (matematyka w zakresie rozszerzonym), $N = 41$: uczniowie klas trzecich tych samych placówek, realizujący program rozszerzony.
- Studenci, $N = 268$: studenci I i II roku kierunków ekonomiczno-menedżerskich dwóch publicznych uczelni w Polsce centralnej, którzy pomyślnie ukończyli akademicki kurs statystyki.

Dobór uczniów klas przedmaturalnych był celowym zabiegiem metodologicznym: na tym etapie nie realizowali oni jeszcze szkolnego modułu statystyki i rachunku prawdopodobieństwa. Pozwoliło to na wyizolowanie ich naturalnej sprawności analitycznej, tworząc punkt odniesienia dla studentów posiadających sformalizowaną wiedzę akademicką. Warto zaznaczyć, że liczebność grupy realizującej program rozszerzony ($N = 41$) odzwierciedla rzeczywistą proporcję uczniów wybierających matematykę rozszerzoną w badanych placówkach. Nierównoliczność grup uwzględniono za pomocą testu t-Welcha.

Sesje badawcze przeprowadzono w kontrolowanych warunkach audytoryjnych techniką ankiety papierowej (PAPI), co wyeliminowało użycie

kalkulatorów, narzędzi AI czy pomocy zewnętrznych. Udział był dobrowolny i anonimowy. Aby zminimalizować presję czasową i ułatwić aktywację Systemu 2, czas odpowiedzi nie był ograniczany. Instrukcja celowo nie sugerowała matematycznego charakteru zadań, zapobiegając sztucznemu wywołaniu podwyższonej czujności poznawczej.

Z uwagi na specyfikę celów badawczych, w pracy zastosowano nielosowy dobór celowy, ukierunkowany na wyizolowanie grup respondentów o ściśle zdefiniowanych profilach edukacyjnych. Choć nielosowy charakter próby nakłada pewne ograniczenia w zakresie bezpośredniej generalizacji wyników na całą populację, zastosowanie aparatu statystyki indukcyjnej (testów chi-kwadrat (χ^2) oraz t-Welcha) posiada silne uzasadnienie metodologiczne. Testy istotności służą jako formalne kryterium oceny, czy zaobserwowane dysproporcje między grupami mają charakter systematyczny, czy też są dziełem przypadku wewnątrz badanej próby. Dodatkowo, w celu uniezależnienia wnioskowania od specyfiki rekrutacji próby, analizę uzupełniono o badanie wielkości efektu (współczynniki V Craméra), co pozwala na rzetelną ocenę rzeczywistej siły zidentyfikowanych zależności. W konsekwencji wyniki testów istotności należy interpretować jako opis relacji zaobserwowanych w obrębie badanej próby, nie zaś jako podstawę do bezpośredniej generalizacji na całą populację uczniów i studentów w Polsce.

Dla przeprowadzonych testów podano wartości statystyk testowych oraz odpowiadające im wartości p (p-value).

Konstrukcja i charakterystyka narzędzia badawczego

Narzędzie badawcze stanowił autorski test sytuacyjny obejmujący cztery zamknięte zadania jednokrotnego wyboru, osadzone w realistycznych kontekstach społeczno-gospodarczych. Test nie miał charakteru sprawdzianu wiedzy szkolnej, lecz został zaprojektowany jako narzędzie diagnostyczne służące ocenie wybranych komponentów alfabetyzmu statystycznego ujawniających się w sytuacjach wymagających interpretacji danych, oceny ryzyka oraz rozumowania probabilistycznego.

Konstrukcja narzędzia została oparta na klasycznych problemach opisywanych w literaturze dotyczącej alfabetyzmu statystycznego, rozumowania statystycznego oraz psychologii poznawczej [Gal 2002; Ben-Zvi i Garfield 2004; Kahneman 2011]. Poszczególne zadania odwoływały się do czterech opisanych w literaturze wzorców błędnego rozumowania: prawa małych liczb, złudzenia gracza, niewrażliwości na zmienność danych oraz błędu proporcji podstawowej, szczegółowo omówionych w części Przeglądu literatury poświęconej pułapkom myślenia w rozumowaniu probabilistycznym. Zmieniono kontekst sytuacyjny, osadzając zadania w realiach współczesnej gospodarki i życia społecznego, co miało zwiększyć ich komunikatywność oraz ograniczyć możliwość odwoływania się przez respondentów do zapamiętanych schematów szkolnych. Każde zadanie reprezentowało odmienny komponent alfabetyzmu statystycznego.

Poprawne rozwiązanie każdego zadania wymagało analitycznej interpretacji informacji statystycznych, charakterystycznej dla Systemu 2, natomiast odpowiedzi niepoprawne (dystraktory) skonstruowano tak, aby odpowiadały wzorcom intuicyjnego rozumowania wiążanym w literaturze z działaniem Systemu 1. Wybór dystraktora wskazywał zatem na odpowiedź zgodną z określonym wzorcem błędnego rozumowania, nie stanowił jednak bezpośredniego pomiaru zastosowania konkretnej heurystyki.

Przed rozpoczęciem badania właściwego narzędzie poddano pilotażowi z udziałem nauczyciela matematyki oraz pięciu uczniów szkoły ponadpodstawowej. Celem pilotażu była ocena zrozumiałości poleceń, jednoznaczności odpowiedzi oraz poprawności konstrukcji zadań. Zgłoszone uwagi posłużyły do doprecyzowania treści pytań i instrukcji. Ogólną strukturę logiczną narzędzia przedstawiono w Tabeli 1.

Tabela 1. Specyfikacja logiczno-strukturalna zadań testowych

Zadanie	Badany komponent alfabetyzmu statystycznego	Dominująca pułapka poznawcza (System 1)	Kompetencja wymagana do poprawnego rozwiązania (System 2)
Z.1: Influencerzy (Zasięgi a proporcja polubień)	Wpływ liczebności próby na zmienność wyników	Prawo małych liczb (heurystyka reprezentatywności)	Ocena wpływu wielkości próby na zmienność wyników
Z.2: Logistyka (Sekwencje tirów przy bramkach)	Rozumienie losowości i niezależności zdarzeń	Złudzenie gracza	Rozpoznanie niezależności kolejnych zdarzeń losowych
Z.3: Zarządzanie (Ocena efektywności pracowników)	Interpretacja zróżnicowania danych	Pomijanie zmienności (koncentracja wyłącznie na średniej)	Wykorzystanie miar rozproszenia do oceny stabilności wyników
Z.4: Systemy AI (Wykrywanie podróbek ofert)	Interpretacja prawdopodobieństwa warunkowego	Błąd proporcji podstawowej	Uwzględnienie częstości bazowej przy ocenie prawdopodobieństwa

Źródło: opracowanie własne

WYNIKI BADAŃ

Analiza wyników skrajnych – podatność i odporność na błędy poznawcze

Pierwszym etapem analizy empirycznej była ocena wyników skrajnych (0 lub 4 punkty). Wynik 0 punktów interpretowano jako pełną podatność na analizowane błędy poznawcze, natomiast wynik 4 punktów jako pełną odporność na te błędy, rozumianą jako zdolność do konsekwentnego odrzucania intuicyjnych, lecz błędnych odpowiedzi i stosowania poprawnego rozumowania analitycznego we wszystkich zadaniach testowych. Wyniki pośrednie (1–3 pkt) poddano odrębnej analizie w kolejnych częściach pracy.

Rozkład wyników skrajnych w poszczególnych grupach badawczych przedstawiono w Tabeli 2.

Tabela 2. Zestawienie wyników skrajnych (0 i 4 punkty) w zależności od profilu edukacyjnego badanych

Grupa badawcza	N	Pełna podatność na analizowane błędy poznawcze (0 pkt)	Pełna odporność na analizowane błędy poznawcze (4 pkt)
Uczniowie (matematyka w zakresie podstawowym)	123	15 (12,2%)	3 (2,4%)
Uczniowie (matematyka w zakresie rozszerzonym)	41	0 (0,0%)	23 (56,1%)
Studenci kierunków ekonomiczno-menedżerskich	268	48 (17,9%)	23 (8,6%)

Objaśnienia: procenty w nawiasach odnoszą się do całkowitej liczebności danej grupy.

Źródło: opracowanie własne na podstawie badań empirycznych

Analiza rozkładu odpowiedzi ujawnia głębokie zróżnicowanie badanych grup:

- Niemal powszechna odporność uczniów z programem rozszerzonym – w tej grupie nie odnotowano ani jednego przypadku całkowitej uległości wobec dystraktorów (0,0%), a ponad połowa respondentów (56,1%) rozwiązała bezbłędnie wszystkie zadania.
- Wyraźna podatność studentów kierunków ekonomiczno-menedżerskich – odsetek osób osiągających wynik zerowy (17,9%) okazał się najwyższy spośród badanych grup i ponad dwukrotnie przewyższał odsetek wyników maksymalnych

(8,6%), co czyni tę grupę najbardziej podatną na intuicyjne pułapki myślenia spośród wszystkich trzech.

- Uczniowie z programem podstawowym bliżej studentów niż rówieśników z rozszerzeniem – niewielki odsetek wyników maksymalnych (2,4%) współwystępował z relatywnie wysokim udziałem wyników zerowych (12,2%).

Istotność statystyczną zależności między profilem edukacyjnym respondentów a poziomem podatności i odporności na błędy poznawcze zweryfikowano testem chi-kwadrat (χ^2). Założenia reguły Cochrańa zostały spełnione (minimalna liczebność oczekiwana wyniosła 7,88), co pozwoliło na prawidłowe przeprowadzenie testu pomimo zerowej liczebności w jednej z komórek tabeli. Analiza wykazała istotny związek między profilem edukacyjnym respondentów a występowaniem wyników skrajnych ($\chi^2(2) = 38,66$; $p < 0,001$). W celu oceny siły tej zależności obliczono współczynnik V Craméra ($V = 0,59$), którego wartość – zgodnie z kryteriami interpretacji efektu zaproponowanymi przez Cohena [1988] – wskazuje na dużą siłę związku. Te same kryteria interpretacji zastosowano konsekwentnie również w dalszej części pracy podczas analizy poszczególnych zadań.

Wyniki te dostarczają pierwszej empirycznej przesłanki odnoszącej się do pytania badawczego PB3: wyższy poziom formalnej edukacji matematyczno-statystycznej nie przekłada się automatycznie na większą odporność na analizowane błędy poznawcze. Co więcej, sam kierunek tej zależności bywa odwrotny do intuicyjnie oczekiwanego – zjawisko to wymaga pogłębionej analizy przedstawionej w dalszej części artykułu.

Rozkład wyboru dystraktorów wskazuje, że część odpowiedzi była zgodna z intuicyjnymi wzorcami przetwarzania informacji charakterystycznymi dla Systemu 1, podczas gdy poprawne rozwiązanie wymagało analitycznego rozumowania charakterystycznego dla Systemu 2. Ze względu na konstrukcję narzędzia nie można jednak utożsamiać wyboru konkretnego dystraktora z bezpośrednim pomiarem zastosowania określonej heurystyki.

Ogólny poziom alfabetyzmu statystycznego – analiza wyników średnich

Kolejnym etapem analizy była ocena ogólnego poziomu analizowanej kompetencji w badanych grupach, wyrażonego średnią liczbą poprawnych odpowiedzi. Statystyki opisowe, przedstawione w Tabeli 3, ujawniają wyraźną hierarchię wyników: najwyższy poziom kompetencji w badanym zakresie i największą jednorodność osiągnęli uczniowie realizujący matematykę w zakresie rozszerzonym. Odmienny obraz uzyskano wśród studentów kierunków ekonomiczno-menedżerskich, którzy osiągnęli średni wynik zbliżony do uczniów realizujących matematykę w zakresie podstawowym, przy jednocześnie największym rozproszeniu wyników.

Tabela 3. Statystyki opisowe wyników testu w badanych grupach

Grupa badawcza	Średnia	Odchylenie standardowe	Współczynnik zmienności
Uczniowie (matematyka w zakresie podstawowym)	1,46	0,86	58,9%
Uczniowie (matematyka w zakresie rozszerzonym)	3,12	1,17	37,4%
Studenci kierunków ekonomiczno-menedżerskich	1,62	1,18	72,6%

Źródło: opracowanie własne na podstawie badań empirycznych

Test Shapiro–Wilka wykazał odstępstwa od rozkładu normalnego we wszystkich grupach: uczniów realizujących matematykę w zakresie podstawowym ($W = 0,864$; $p < 0,001$), uczniów realizujących matematykę w zakresie rozszerzonym ($W = 0,722$; $p < 0,001$) oraz studentów ($W = 0,900$; $p < 0,001$). Test Browna–Forsythe’a wskazał ponadto na brak homogeniczności wariancji ($F(2, 429) = 4,06$; $p = 0,018$). Ze względu na nierówne liczebności grup i zróżnicowanie wariancji do porównania średnich zastosowano dwustronny test t-Welcha dla prób niezależnych. Przy liczebnościach analizowanych grup test ten pozostaje dostatecznie odporny na odstępstwa od normalności; stabilność uzyskanych wniosków zweryfikowano dodatkowo testem U Manna–Whitneya. Ponieważ przeprowadzono trzy porównania, zastosowano poprawkę Bonferroniego, przyjmując skorygowany poziom istotności $\alpha = 0,0167$.

Porównania parami średnich wyników ujawniają wyraźny wzorec:

- Rozszerzony program wyprzedza podstawowy – różnica średnich okazała się istotna statystycznie ($t(55,29) = 8,38$; $p < 0,001$), na korzyść uczniów realizujących matematykę w zakresie rozszerzonym.
- Rozszerzony program wyprzedza też studia – stwierdzono istotną statystycznie różnicę ($t(53,22) = 7,68$; $p < 0,001$), również na korzyść uczniów z programem rozszerzonym.
- Studenci nie odróżniają się od uczniów z programem podstawowym – nie stwierdzono istotnej statystycznie różnicy średnich ($t(314,64) = 1,47$; $p = 0,141$). Wartość p przekroczyła zarówno klasyczny, jak i skorygowany poziom istotności, co oznacza brak podstaw do odrzucenia hipotezy o braku różnic między średnimi wynikami obu grup. W badanej próbie średni poziom kompetencji studentów po ukończeniu akademickiego kursu statystyki nie różnił się istotnie od poziomu reprezentowanego przez uczniów realizujących matematykę wyłącznie w zakresie podstawowym.

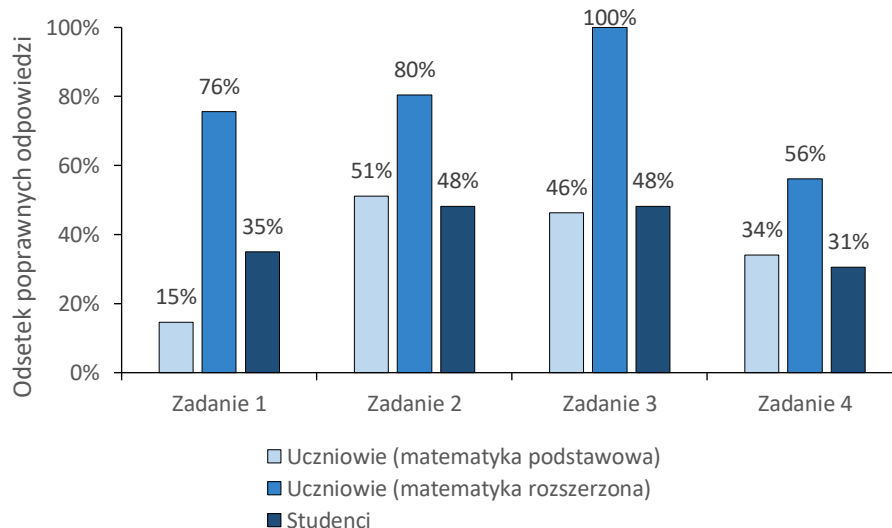
Kontrolne porównania testem U Manna–Whitneya potwierdziły wszystkie rozstrzygnięcia uzyskane testem Welcha: dla uczniów realizujących matematykę w zakresie rozszerzonym i podstawowym uzyskano $U = 802,5$; $p < 0,001$, dla uczniów realizujących matematykę w zakresie rozszerzonym i studentów $U = 8822,5$; $p < 0,001$, natomiast dla uczniów realizujących matematykę w zakresie podstawowym i studentów $U = 15705$; $p = 0,435$.

Uzyskane wyniki stanowią bezpośrednią odpowiedź na pytania badawcze PB1 i PB3. Analiza wykazała istotne zróżnicowanie ogólnego poziomu alfabetyzmu statystycznego pomiędzy badanymi grupami oraz wskazała, że wzrost poziomu formalnej edukacji matematyczno-statystycznej nie prowadzi automatycznie do wyższych rezultatów w tym zakresie. Świadczy o tym fakt, że studenci, pomimo ukończenia akademickiego kursu statystyki, nie uzyskali wyników lepszych od uczniów realizujących matematykę w zakresie rozszerzonym – to właśnie ta grupa osiągnęła najwyższy poziom analizowanej kompetencji. Anomalia ta wskazuje, że sam wzrost poziomu formalnej edukacji matematyczno-statystycznej nie gwarantuje ani wyższego poziomu alfabetyzmu statystycznego, ani większej odporności na analizowane błędy poznawcze. Możliwe przyczyny tego zjawiska omówiono w części poświęconej dyskusji wyników.

Analiza poszczególnych komponentów alfabetyzmu statystycznego

W kolejnym etapie analizy oceniono odsetek poprawnych odpowiedzi uzyskanych w poszczególnych zadaniach testowych. Pozwoliło to ustalić, które obszary alfabetyzmu statystycznego sprawiały respondentom największe trudności oraz w jakim stopniu wyniki różniły się pomiędzy badanymi grupami. Odsetek poprawnych odpowiedzi w podziale na profil edukacyjny badanych przedstawiono na Rysunku 1.

Rysunek 1. Odsetek poprawnych odpowiedzi w poszczególnych zadaniach testowych z podziałem na profil edukacyjny badanych



Źródło: opracowanie własne

Zależność pomiędzy profilem kształcenia a poprawnością odpowiedzi w poszczególnych zadaniach testowych zweryfikowano za pomocą testu chi-kwadrat (χ^2). Każde z analizowanych zadań zostało zaprojektowane w celu oceny innego komponentu, przedstawionego w części dotyczącej konstrukcji narzędzia badawczego. Ze względu na przeprowadzenie czterech niezależnych testów na tym samym zbiorze danych zastosowano poprawkę Bonferroniego, przyjmując skorygowany poziom istotności $\alpha = 0,0125$. Wszystkie uzyskane wartości p okazały się niższe od tego progu, co potwierdza istotne statystycznie zróżnicowanie poprawności odpowiedzi pomiędzy badanymi grupami.

Cztery zadania testowe ujawniają odmienne oblicza tej samej prawidłowości. Dla każdego zadania przedstawiono wartość statystyki χ^2 , odpowiadającą jej wartość p oraz współczynnik V Craméra jako miarę siły związku:

- Wpływ wielkości próby (Zadanie 1) – najwyższy odsetek poprawnych odpowiedzi uzyskali uczniowie realizujący matematykę w zakresie rozszerzonym (76%). Niższe wyniki odnotowano wśród studentów (35%) oraz uczniów realizujących matematykę w zakresie podstawowym (15%). Analiza wykazała istotny statystycznie związek między profilem edukacyjnym respondentów a poprawnością odpowiedzi ($\chi^2(2) = 52,87$; $p < 0,001$), a wartość współczynnika V Craméra ($V = 0,35$) wskazuje na umiarkowaną siłę tego związku.
- Rozumienie losowości (Zadanie 2) – 80% uczniów realizujących matematykę w zakresie rozszerzonym poprawnie rozpoznało niezależność zdarzeń, podczas

gdy wyniki pozostałych dwóch grup oscylowały wokół poziomu losowego zgadywania (51% i 48%). Rozkład poprawnych odpowiedzi pozostawał w istotnym statystycznie związku z profilem edukacyjnym respondentów ($\chi^2(2) = 14,97$; $p = 0,0006$). Wartość współczynnika V Craméra ($V = 0,19$) wskazuje jednak na słabą siłę tego związku.

- Interpretacja zmienności danych (Zadanie 3) – wszyscy uczniowie realizujący matematykę w zakresie rozszerzonym rozwiązyli zadanie poprawnie (100%), podczas gdy studenci oraz uczniowie realizujący matematykę w zakresie podstawowym uzyskali wyniki ponad dwukrotnie niższe (odpowiednio 48% i 46%). Zależność między profilem edukacyjnym respondentów a poprawnością odpowiedzi okazała się istotna statystycznie ($\chi^2(2) = 41,02$; $p < 0,001$), a wartość współczynnika V Craméra ($V = 0,31$) wskazuje na umiarkowaną siłę tego związku. Warto zauważyć, że jest to jedyne z czterech zadań, w którym wszyscy uczniowie realizujący program rozszerzony udzielili poprawnej odpowiedzi.
- Prawdopodobieństwo warunkowe (Zadanie 4) – komponent ten okazał się najtrudniejszy dla uczniów realizujących matematykę w zakresie rozszerzonym oraz dla studentów, którzy uzyskali tu swoje najniższe wyniki spośród wszystkich czterech zadań (odpowiednio 56% i 31%). Wyjątek stanowili uczniowie realizujący matematykę w zakresie podstawowym, dla których niższy wynik odnotowano jedynie w Zadaniu 1 (15%); w Zadaniu 4 uzyskali oni 34% poprawnych odpowiedzi. Poprawność odpowiedzi pozostawała w istotnym statystycznie związku z profilem edukacyjnym respondentów ($\chi^2(2) = 10,30$; $p = 0,0058$). Wartość współczynnika V Craméra ($V = 0,15$) wskazuje jednak na słabą siłę tego związku.

Analiza poszczególnych komponentów badanej kompetencji pozwoliła odpowiedzieć na pytanie badawcze PB2. Największe trudności respondentom sprawiało prawdopodobieństwo warunkowe – zwłaszcza uczniom z programem rozszerzonym oraz studentom – natomiast dla uczniów z programem podstawowym największym wyzwaniem okazało się rozumienie wpływu wielkości próby. We wszystkich analizowanych obszarach najwyższe odsetki poprawnych odpowiedzi uzyskali uczniowie realizujący matematykę w zakresie rozszerzonym. Wyniki te uszczegóławiają rezultaty przedstawione w części poświęconej porównaniu średnich wyników grup. Analiza siły zależności miarą V Craméra ujawniła, że za zróżnicowanie poziomu kompetencji pomiędzy badanymi grupami w największym stopniu odpowiadały komponenty związane z wielkością próby (Zadanie 1) oraz interpretacją zmienności danych (Zadanie 3), gdzie odnotowano umiarkowaną siłę związku.

Podsumowując, analiza poszczególnych komponentów potwierdziła, że wzrost poziomu formalnej edukacji matematyczno-statystycznej nie zawsze przekłada się na wyższy poziom badanej kompetencji. Obserwacja ta jest zgodna z przyjętym w niniejszym artykule rozumieniem luki kompetencyjnej jako rozbieżności pomiędzy efektami formalnej edukacji matematyczno-statystycznej

a poziomem alfabetyzmu statystycznego ujawnianym podczas interpretacji danych i rozumowania probabilistycznego.

WNIOSKI I DYSKUSJA

Synteza wyników

Przeprowadzone badanie wykazało trzy najważniejsze prawidłowości:

- Zróżnicowanie kompetencyjne między grupami – poziom alfabetyzmu statystycznego istotnie różnił się pomiędzy analizowanymi grupami, przy czym najwyższe wyniki uzyskali uczniowie realizujący matematykę w zakresie rozszerzonym.
- Zróżnicowanie trudności poszczególnych komponentów kompetencji – uczniom realizującym matematykę w zakresie rozszerzonym oraz studentom największą trudność sprawiało prawdopodobieństwo warunkowe, natomiast w grupie uczniów realizujących matematykę w zakresie podstawowym najniższy wynik odnotowano w zadaniu dotyczącym wpływu wielkości próby.
- Brak jednokierunkowego wzrostu kompetencji wraz z poziomem edukacji – studenci, mimo ukończenia akademickiego kursu statystyki, nie uzyskali statystycznie istotnej przewagi nad uczniami realizującymi matematykę w zakresie podstawowym. Najwyższy poziom badanej kompetencji stwierdzono natomiast wśród uczniów realizujących matematykę w zakresie rozszerzonym.

Dyskusja wyników

Przeprowadzone badanie ujawniło w analizowanej próbie swoistą anomalię edukacyjną: wyniki nie układały się zgodnie z oczekiwanym, jednokierunkowym wzrostem alfabetyzmu statystycznego w grupach reprezentujących kolejne etapy formalnej edukacji. Prawidłowość ta dotyczyła również odporności na analizowane błędy poznawcze. Wyniki nie wskazują zatem, aby osoby po ukończeniu kursu statystyki były w badanej próbie bardziej odporne na uleganie heurystykom. Podobny brak jednorodności rozwoju alfabetyzmu statystycznego pomiędzy kolejnymi poziomami kształcenia odnotowali również Kurnia, Lowrie i Patahuddin [2024], co wskazuje, że stwierdzona anomalia nie stanowi odosobnionego przypadku w literaturze przedmiotu.

Jednym z możliwych wyjaśnień uzyskanego układu wyników jest odmienny charakter kompetencji rozwijanych w toku edukacji szkolnej i akademickiej. Nauczanie statystyki na poziomie akademickim koncentruje się często na opanowaniu metod obliczeniowych i procedur statystycznych, podczas gdy zadania wykorzystane w niniejszym badaniu wymagały przede wszystkim interpretacji informacji, krytycznej oceny danych oraz świadomego wnioskowania w warunkach niepewności. Z kolei systematyczne rozwiązywanie w szkole średniej zadań wymagających uzasadniania i weryfikowania własnego rozumowania może sprzyjać gotowości do podważania pierwszych intuicji. Na stwierdzone różnice mogły wpływać również czynniki selekcyjne, w tym wcześniejsze kompetencje

matematyczne i motywacja uczniów wybierających zakres rozszerzony, których wpływu nie można oddzielić od wpływu sposobu kształcenia w badaniu przekrojowym. Interpretacja ta ma jednak charakter hipotetyczny i wymaga weryfikacji w dalszych badaniach.

Uzyskane rezultaty wpisują się w opisywaną w literaturze rozbieżność pomiędzy opanowaniem procedur statystycznych a umiejętnością ich praktycznego wykorzystania. Są również zgodne z koncepcją alfabetyzmu statystycznego przedstawioną przez Gała [2002] oraz Ben-Zviego i Garfielda [2004], zgodnie z którą poprawne rozumowanie statystyczne wykracza poza znajomość metod obliczeniowych i obejmuje interpretację informacji, krytyczną ocenę danych oraz podejmowanie decyzji w warunkach niepewności. Wyniki badania sugerują zatem, że kompetencje rachunkowe i interpretacyjne mogą rozwijać się częściowo niezależnie, co stanowi istotne wyzwanie dla dydaktyki statystyki.

Implikacje dydaktyczne

Uzyskane wyniki wskazują na potrzebę wzmacniania w kształceniu statystycznym umiejętności interpretowania danych i świadomego rozumowania w sytuacjach decyzyjnych. W szczególności zasadne wydają się:

- integrowanie treści statystycznych z podstawami ekonomii behawioralnej w celu uświadamiania studentom mechanizmów działania heurystyk poznawczych opisanych w literaturze [Tversky i Kahneman 1974];
- zwiększenie udziału zadań opartych na rzeczywistych danych i autentycznych problemach decyzyjnych oraz osadzonych w realistycznym kontekście gospodarczym;
- większe akcentowanie interpretacji wyników statystycznych i oceny wiarygodności wniosków, obok rozwijania umiejętności wykonywania obliczeń;
- wykorzystywanie zadań wymagających uzasadniania odpowiedzi i analizowania alternatywnych sposobów rozumowania, co może sprzyjać rozwijaniu alfabetyzmu statystycznego.

Ograniczenia badania

Interpretując uzyskane wyniki, należy uwzględnić trzy najważniejsze ograniczenia badania. Po pierwsze, zastosowano celowy dobór próby, co ogranicza możliwość uogólniania rezultatów na populację uczniów i studentów w Polsce oraz nie pozwala formułować wniosków o przyczynowym wpływie kolejnych etapów edukacji na poziom badanej kompetencji.

Po drugie, grupa uczniów realizujących matematykę w zakresie rozszerzonym była mniej liczna niż pozostałe grupy. Pomimo zastosowania metod statystycznych uwzględniających nierównoliczność prób wskazana jest ostrożność w interpretacji wielkości stwierdzonych różnic oraz ich weryfikacja na liczniejszych próbach.

Po trzecie, zastosowane narzędzie obejmowało cztery zadania odnoszące się do wybranych komponentów alfabetyzmu statystycznego. Wyniki dotyczą zatem

analizowanych obszarów kompetencji i nie wyczerpują całego zakresu tego złożonego konstruktu.

Kierunki dalszych badań

Naturalną kontynuacją badania będzie szczegółowa analiza rozkładu wyboru poszczególnych dystraktorów, pozwalająca zidentyfikować najczęstsze schematy błędnego rozumowania oraz lepiej rozpoznać mechanizmy prowadzące do nieprawidłowej interpretacji danych i oceny niepewności. Analiza ta zostanie przedstawiona w odrębnym opracowaniu poświęconym jakościowej interpretacji błędów poznawczych ujawnianych przez respondentów. Celowe wydaje się również rozszerzenie badań na studentów innych kierunków i osoby o zróżnicowanym doświadczeniu zawodowym, a także dalsze doskonalenie i walidacja zaproponowanego narzędzia badawczego.

Podsumowanie

Rozwój alfabetyzmu statystycznego wymaga nie tylko opanowania metod i narzędzi ilościowych, lecz także systematycznego kształtowania umiejętności interpretowania danych oraz poprawnego rozumowania w warunkach niepewności. Skuteczność kształcenia matematyczno-statystycznego warto zatem oceniać zarówno przez pryzmat poprawności wykonywanych obliczeń, jak i zdolności do świadomego i krytycznego wykorzystywania informacji statystycznych w procesie podejmowania decyzji.

BIBLIOGRAFIA

- Bar-Hillel M. (1980) The Base-Rate Fallacy in Probability Judgments. *Acta Psychologica*, 44(3), 211-233. [https://doi.org/10.1016/0001-6918\(80\)90046-3](https://doi.org/10.1016/0001-6918(80)90046-3)
- Batanero C., Chernoff E. J., Engel J., Lee H. S., Sánchez E. (2016) Research on Teaching and Learning Probability. ICME-13 Topical Surveys. Springer, Cham. <https://doi.org/10.1007/978-3-319-31625-3>
- Ben-Zvi D., Garfield J. (Red.) (2004) The Challenge of Developing Statistical Literacy, Reasoning and Thinking. Springer Science & Business Media. <https://doi.org/10.1007/1-4020-2278-6>
- Cohen J. (1988) Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Domański C. (2016) Wyzwania w zakresie edukacji statystycznej społeczeństwa. *Wiadomości Statystyczne. The Polish Statistician*, 9(664), 105-115. <https://doi.org/10.5604/01.3001.0014.1099>
- Gal I. (2002) Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review*, 70(1), 1-25. <https://doi.org/10.1111/j.1751-5823.2002.tb00336.x>
- Gigerenzer G., Hoffrage U. (1995) How to Improve Bayesian Reasoning without Instruction: Frequency Formats. *Psychological Review*, 102(4), 684-704. <https://doi.org/10.1037/0033-295X.102.4.684>

- Gould R. (2017) Data Literacy is Statistical Literacy. *Statistics Education Research Journal Archives*, 16(1), 22-25. <https://doi.org/10.52041/serj.v16i1.209>
- Jasiewicz J., Stopnicka M. (2016) Data literacy: kompetencje w zakresie korzystania z danych w odniesieniu do przemian komunikacji naukowej. *Kultura Popularna*, 50(1), 36-49. <https://doi.org/10.5604/01.3001.0010.0042>
- Kahneman D. (2011) *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
- Kahneman D., Tversky A. (1972) Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology*, 3(3), 430-454. [https://doi.org/10.1016/0010-0285\(72\)90016-3](https://doi.org/10.1016/0010-0285(72)90016-3)
- Kurnia A. B., Lowrie T., Patahuddin S. M. (2024) The Development of High School Students' Statistical Literacy across Grade Level. *Mathematics Education Research Journal*, 36(Suppl 1), 7-35. <https://doi.org/10.1007/s13394-023-00449-x>
- Oleński J. (2020) Środowisko informacyjne statystyki publicznej. *Wiadomości Statystyczne. The Polish Statistician*, 65(4), 32-45. <https://doi.org/10.5604/01.3001.0014.1229>
- Phadke S., Beckman M., Lock Morgan K. (2024) Examining the Role of Context in Statistical Literacy Assessment. *Statistics Education Research Journal*, 23(1), Article 4. <https://doi.org/10.52041/serj.v23i1.529>
- Szreder M. (2016) O niektórych nowych wyzwaniach i oczekiwaniach wobec statystyki. *Wiadomości Statystyczne. The Polish Statistician*, 61(6), 1-9. <https://doi.org/10.5604/01.3001.0014.1000>
- Tversky A., Kahneman D. (1971) Belief in the Law of Small Numbers. *Psychological Bulletin*, 76(2), 105-110. <https://doi.org/10.1037/h0031322>
- Tversky A., Kahneman D. (1974) Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-1131. <https://doi.org/10.1126/science.185.4157.1124>
- Wild C. J., Pfannkuch M. (1999) Statistical Thinking in Empirical Enquiry. *International Statistical Review*, 67(3), 223-248. <https://doi.org/10.1111/j.1751-5823.1999.tb00442.x>

STATISTICAL LITERACY AND THINKING TRAPS: AN ANALYSIS OF THE COMPETENCY GAP AT THE TRANSITION FROM SECONDARY TO HIGHER EDUCATION

Abstract: The aim of this paper is to assess the level of statistical literacy among secondary school and university students and to analyse selected components of this competence. The study was conducted on a sample of 432 secondary school and university students, using an original scenario-based test. The findings indicate the existence of a competency gap between the outcomes of formal mathematics and statistics education and the level of statistical literacy demonstrated in tasks involving data interpretation and probabilistic reasoning. The analysis revealed an educational anomaly: students completing the extended mathematics curriculum achieved a higher level of statistical literacy than university students who had completed an academic statistics course.

Keywords: statistical literacy, competency gap, statistics education, cognitive heuristics, probabilistic reasoning, data interpretation

JEL classification: A21, A22, D91, C12, I21

MITIGATING CLASS IMBALANCE IN U.S. FINANCIAL SECURITY: A LATENT FEATURE AUGMENTATION APPROACH FOR HIGH-PURITY FRAUD DETECTION

Kaniz Sultana Chy  <https://orcid.org/0009-0002-9924-0081>

Dept. of Information Systems, Lamar University, Beaumont, Texas, USA
e-mail: kanizsultanachy@gmail.com

Md Shoriful Islam Chowdhury  <https://orcid.org/0009-0006-1179-4901>

Department of Public Administration, University of Chittagong, Chattogram, BD
e-mail: shorifuli676@gmail.com

Md Nurul Islam Chowdhury  <https://orcid.org/0009-0005-6482-621X>

Department of Public Administration, University of Chittagong, Chattogram, BD
e-mail: bipschy@gmail.com

Md Ashiqul Islam  <https://orcid.org/0009-0002-1505-9446>

College of GPS, Trine University, Angola, Indiana, USA
e-mail: ashiquislam3272@gmail.com

Md Asiful Islam  <https://orcid.org/0009-0006-9913-9433>

Department of Business Administration, Green University, Narayanganj, BD
e-mail: asifulapu@gmail.com

Abstract: Credit card fraud detection is challenged by extreme class imbalance ($< 0.2\%$ fraud) and non-linear fraud patterns, causing traditional supervised models to suffer from reduced sensitivity and poor interpretability. This paper proposes a hybrid model combining a sparse Deep Autoencoder (15-node bottleneck) with Random Forest classification via feature augmentation—concatenating raw features with latent representations and reconstruction errors. Validation on 85,442 transactions yields PR-AUC of 0.6942, precision of 0.9412 ± 0.0318 , and recall of 0.7968 ± 0.0354 across 5-fold cross-validation. While performance is statistically comparable to standard Random Forest ($p > 0.05$), the hybrid model offers enhanced interpretability: the Anomaly Score ranks as the second most important predictor, and t-SNE reveals distinct geometric separation of fraud clusters. These findings demonstrate that unsupervised feature augmentation provides a viable, interpretable solution for real-time fraud monitoring.

Keywords: Hybrid Manifold Learning, Credit Card Fraud Detection, Sparse Autoencoders, Class Imbalance Mitigation, Precision-Recall Optimization

JEL classification: C45, C55, C63

<https://doi.org/10.22630/MIBE.2026.27.3.13>



INTRODUCTION

The rapid growth of electronic transactions has increased global exposure to sophisticated financial fraud, costing billions annually [Utami et al. 2020]. Traditional rule-based systems struggle against dynamic, non-linear fraud patterns [Vadapalli et al. 2026], while computational models face extreme class imbalance (<0.2% fraud) and high-dimensional noise [Chougule, Mane 2026; Xie et al. 2023]. Supervised algorithms often skew toward majority normal classes [Chougule, Mane 2026], while raw features induce overfitting [Xie et al. 2023]. Conversely, unsupervised autoencoders effectively map normal behavior and generate anomaly scores [Alhabib et al. 2023], yet lack the discriminative strength of tabular supervised ensembles like Random Forests [6].

Recent multi-stage architectures have sought to address these constraints. Lin & Jiang [2021] combined Autoencoders with Probabilistic Random Forests (AE-PRF) to eliminate explicit resampling [Chen et al. 2018]. Subsequent studies underscored that ensemble methods require precise feature reduction to mitigate noise sensitivity [Kar & Dasgupta 2024], while temporal models (LSTM/GRU-Autoencoders) often suffer from information loss by replacing raw data with compressed embeddings [Sun et al. 2025; Ezhilmathi & Samy 2026]. Furthermore, a persistent interpretability gap hinders regulatory compliance (e.g., GDPR, PCI-DSS) by failing to provide geometric transparency for flagged transactions [Bin Sulaiman et al. 2022; El Alami et al. 2026]. Various resampling techniques, such as SMOTE (Synthetic Minority Over-sampling Technique), have been proposed to address class imbalance by generating synthetic fraud samples by Mallik et al [2019]. However, these methods can introduce noise and overfitting risks when applied to high-dimensional transaction data from Chawla et al. [2002]. The challenges of learning from imbalanced data have been extensively surveyed by He & Garcia [2009], who identified that standard classifiers often fail to identify minority class patterns, particularly in high-stakes domains such as fraud detection.

To bridge this information loss gap, this study proposes a Feature Augmentation strategy. Rather than discarding original attributes [as in Lin & Jiang 2021; Chen et al. 2018], our framework retains raw features alongside a 15-node sparse latent manifold (L_2 and KL regularization) and reconstruction residuals, allowing a Random Forest to evaluate raw and unsupervised evidence simultaneously. Validated on 85,442 samples via 5-fold cross-validation, the model achieves strong performance (Precision: 0.9412 ± 0.0318 ; Recall: 0.7968 ± 0.0354 ; $p > 0.05$) while significantly improving interpretability—the reconstruction error emerged as the second most vital predictor, and t-SNE projections confirmed clear geometric separation of fraudulent instances in the latent manifold.

PROBLEM FORMULATION AND METHODOLOGY

Problem Statement

Consider a transaction dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$, where $x_i \in R^{30}$ represents a feature vector and $y_i \in \{0,1\}$ is the class label (0 = normal, 1 = fraud). The primary challenge in this domain is extreme class imbalance, where $N_{\text{fraud}} \ll N_{\text{normal}}$ (fraud represents $< 0.2\%$ of all transactions).

This research seeks to develop a hybrid mapping function $\mathcal{H}: R^{30} \rightarrow \{0,1\}$ that maximizes the Precision-Recall Area Under the Curve (PR-AUC) by leveraging unsupervised manifold learning to augment supervised decision boundaries.

Stage I: Unsupervised Manifold Learning (Deep Autoencoder)

The first stage employs a Deep Autoencoder to extract a compressed, non-linear representation of the transaction space. This process filters high-dimensional noise and identifies the underlying structure of normal behavior.

The encoder maps the input vector $x \in R^{30}$ to a latent bottleneck representation $z \in R^{15}$ using a weighted transformation and a log-sigmoid activation function σ :

$$z = \sigma(W_1 x + b_1) \quad (1)$$

Where $\sigma(a) = \frac{1}{1+e^{-a}}$, $W_1 \in R^{15 \times 30}$, and $b_1 \in R^{15}$. The 15-node bottleneck was selected through preliminary sensitivity analysis (Section 3.5) to balance compression with information preservation.

The decoder recovers the reconstructed input space $\hat{x} = \sigma(W_2 z + b_2)$, where $W_2 \in R^{30 \times 15}$ and $b_2 \in R^{30}$. The network optimization minimizes a composite loss function $\mathcal{J}(W, b)$ penalizing reconstruction error, weight complexity (L_2 regularization), and activation density (sparsity):

$$\mathcal{J}(W, b) = \underbrace{\frac{1}{n} \sum_{i=1}^n \frac{1}{2} |x_i - \hat{x}_i|^2}_{\text{MSE Loss}} + \underbrace{\frac{\lambda}{2} \sum_l |W_l|_F^2}_{\text{Weight Decay}} + \underbrace{\beta \sum_{j=1}^{15} KL(\rho | \hat{\rho}_j)}_{\text{Sparsity Penalty}} \quad (2)$$

- Mean Squared Error (MSE): Measures input decoding discrepancy.
- Weight Decay: Controls model complexity with parameter $\lambda = 0.001$.
- Sparsity Penalty: Uses Kullback-Leibler (KL) divergence with target activation $\rho = 0.05$ and penalty coefficient $\beta = 4$.

The KL divergence for hidden unit j over training average activation $\hat{\rho}_j$ is defined as:

$$KL(\rho \parallel \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \quad (3)$$

This formulation forces the network to capture sparse, salient structural properties over raw noise.

Stage II: Supervised Ensemble Classification (Random Forest)

The representations extracted by the autoencoder are transferred to a Random Forest (RF) classifier using an augmented feature space.

The augmented feature vector x' combines raw inputs, bottleneck features, and an anomaly signal:

$$x' = [x, z, \epsilon] \quad (4)$$

Where $\epsilon = \|x - \hat{x}\|_2$ denotes the scalar reconstruction error (Anomaly Score). This allows the ensemble classifier to evaluate raw variables alongside latent structural encodings simultaneously.

The model constructs an ensemble of $B = 100$ decision trees $\{T_b\}_{b=1}^B$. Each tree is trained on a bootstrap sample of x' , with final decisions assigned via majority vote:

$$\hat{y} = \arg \max_{c \in \{0,1\}} \left(\frac{1}{B} \sum_{b=1}^B I(T_b(x') = c) \right) \quad (5)$$

Tree splits optimize Gini Impurity (I_G). To mitigate extreme target skew, an asymmetric cost matrix C penalizes false negatives (missed fraud) ten times higher than false positives (false alarms):

$$C = \begin{bmatrix} 0 & 1 \\ 10 & 0 \end{bmatrix} \quad (6)$$

To validate structural separability within latent space z , t-Distributed Stochastic Neighbor Embedding (t-SNE) maps bottleneck features by minimizing KL divergence between high-dimensional joint probabilities P and low-dimensional embedding probabilities Q :

$$\mathcal{C} = KL(P \parallel Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (7)$$

The Barnes-Hut algorithm variant (perplexity = 30) was evaluated across a stratified sample of 5,000 test set instances.

Experimental Setup

Experiments were evaluated on the Kaggle Credit Card Fraud Detection dataset (284,807 instances, 30 features). Features include 28 anonymized PCA vectors, Time, and Amount. Two domain features were engineered prior to training:

$$\text{Hour Of Day} = \left(\frac{\text{Time}}{3600} \right) \quad (8)$$

$$\text{Normalized Amount} = \frac{\text{Amount} - \mu_{\text{Amount}}}{\sigma_{\text{Amount}}} \quad (9)$$

The full dataset was shuffled and split using a 70/30 stratified train-test scheme. Hyperparameter tuning relied on 5-fold cross-validation across the training set. Table 1 shows the hyperparameter configurations.

Table 1. Hyperparameter configuration for the proposed hybrid pipeline.

Pipeline Component	Parameter Name	Notation	Configured Value
Deep Autoencoder	Bottleneck Hidden Nodes	$ z $	15
	Maximum Training Epochs	—	20
	L_2 Weight Regularization	λ	0.001
	Sparsity Target Activation	ρ	0.05
	Sparsity Penalty Weight	β	4
Random Forest	Total Decision Trees	B	100
	Asymmetric Cost Matrix	C	[0,1; 10,0]
	Out-of-Bag Evaluation	—	Enabled
Experimental Setup	Train/Test Split	—	70 / 30
	Cross-Validation Folds	—	5
	t-SNE Perplexity	—	30
	Random State Seed	—	42

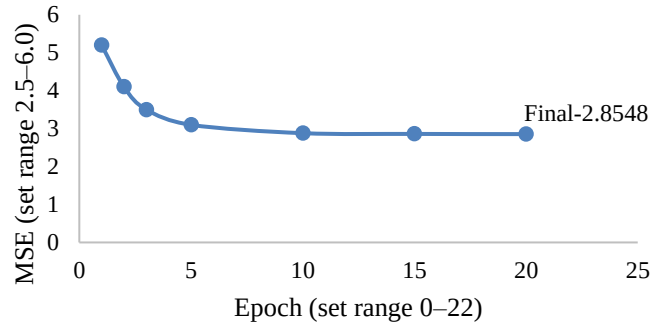
Source: own calculations

Performance metrics across 5-fold cross-validation are reported as mean $\pm$ standard deviation. Statistically significant improvements against baseline models were evaluated via paired two-tailed t-tests ($\alpha = 0.05$). Pipelines were implemented using MATLAB R2018a (Statistics and Machine Learning Toolbox, Deep Learning Toolbox).

RESULTS & DISCUSSIONS

The initial experimental phase evaluated the unsupervised learning capacity of the Deep Autoencoder in characterizing the transactional manifold. As shown in Fig. 1, the training profile demonstrates robust convergence, stabilizing at a Mean Squared Error (MSE) of 2.8548 by the 20th epoch. This rapid stabilization confirms that combining L_2 regularization with sparsity constraints successfully suppressed high-frequency noise, enabling the 15-node bottleneck layer to capture structural representations.

Figure 1: Autoencoder Training Convergence (15-node)



Source: own preparations

To establish statistical validity across model iterations, a 5-fold cross-validation scheme was executed on the training split. Table 2 details the mean precision and recall for both the baseline Random Forest and the proposed Hybrid model across all folds.

Table 2. 5-fold cross-validation results (mean $\pm$ standard deviation)

Model	Precision (mean $\pm$ std)	Recall (mean $\pm$ std)
Standard Random Forest	0.9477 $\pm$ 0.0293	0.7988 $\pm$ 0.0213
Hybrid Random Forest	0.9412 $\pm$ 0.0318	0.7968 $\pm$ 0.0354
p-value (Paired t-test)	0.0789	0.8160

Source: own calculations

A paired t-test indicates no statistically significant performance difference between the architectures (precision $p = 0.0789$; recall $p = 0.8160$). However, the hybrid pipeline delivers structural interpretability by revealing latent fraud signatures while preserving baseline accuracy.

The full 70% training partition was used to fit the final models, which were subsequently benchmarked against the 30% held-out test dataset. Performance metrics are summarized in Table 3.

Table 3. Hold-out test set performance evaluation (70/30 split)

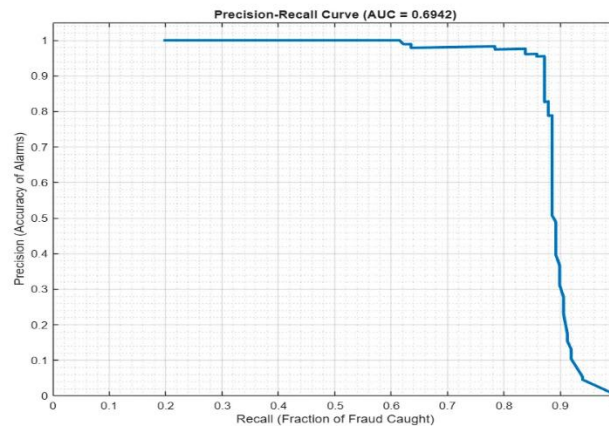
Model	Precision	Recall	F1-Score
Standard Random Forest	0.9542	0.8446	0.8961
Hybrid Random Forest	0.9621	0.8581	0.9071

Source: own calculations

The hybrid architecture achieved a precision of 0.9621 and a recall of 0.8581, yielding an F_1 -score of 0.9071. This balance between false positive reduction and true fraud capture demonstrates suitability for production financial monitors.

Fig. 2 highlights the Precision-Recall (PR) performance curve generated from the test set. The model yields an overall PR-AUC of 0.6942 under extreme class skew ($< 0.2\%$ positive rate).

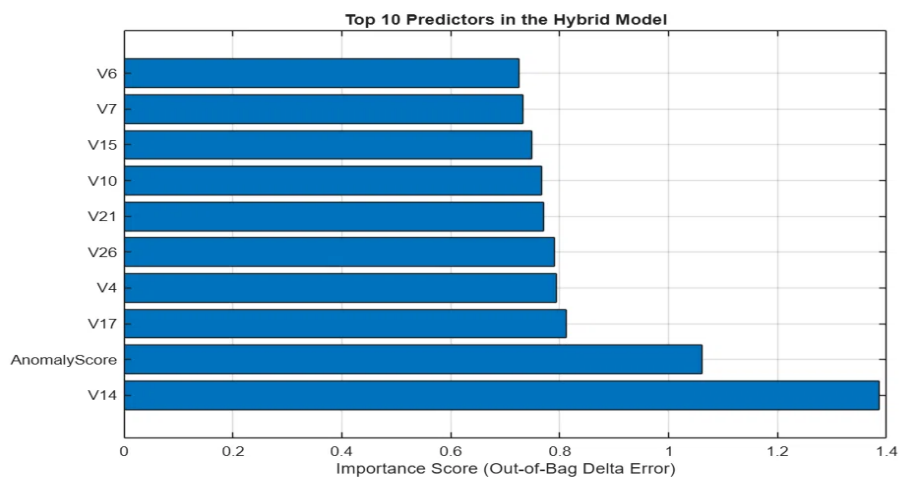
Figure 2. Precision-Recall (PR) curve for hybrid model



Source: own preparations

Precision remains near optimal across conservative operational thresholds, exhibiting a gradual decay only as recall surpasses 0.80. This trajectory highlights effective control over operational false-alarm generation under dense transaction throughput. Feature contributions within the Random Forest ensemble were quantified using Out-of-Bag (OOB) permuted delta error.

Figure 3. Top 10 predictors in Hybrid model

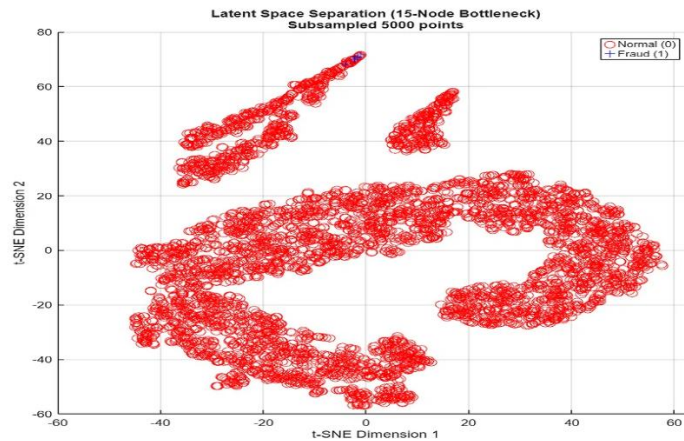


Source: own calculations

Figure 3 shows the autoencoder reconstruction error (Anomaly Score) ranked as the second most influential variable (score: 1.20), trailing only V14 (1.35). This confirms that the unsupervised stage successfully generated non-linear anomaly features that enrich the decision trees.

To assess topological structure, the 15-dimensional bottleneck representation z was projected into a 2D space using Barnes-Hut t-SNE (Fig. 4).

Figure 4. 15-dimensional bottleneck layer into a 2D space using Barnes-Hut t-SNE



Source: own calculations

The geometric isolation of fraudulent transactions within localized high-density islands confirms that the bottleneck network effectively isolates non-linear boundaries separating normal and suspicious behavior. Operationally, this hybrid pipeline delivers significant advantages: the high feature importance of the Anomaly Score provides clear diagnostic signals for forensic auditors, while a high precision of 0.9621 minimizes labor spent reviewing false positives. Furthermore, this structural interpretability supports regulatory compliance standards like GDPR and PCI-DSS, and the low inference latency allows smooth integration into real-time transaction processing systems.

However, several constraints limit these empirical conclusions. Testing was restricted to the PCA-anonymized Kaggle dataset, requiring further validation on raw, multi-source data streams. Additionally, the lack of a statistically significant performance improvement over baseline Random Forest ($p > 0.05$) indicates that gains depend heavily on input feature quality. Finally, deploying this architecture in ultra-low-latency streaming applications may necessitate model quantization or distillation to optimize processing speed.

SUMMARY

This study demonstrates that enriching supervised tree ensembles with unsupervised latent features effectively resolves the dual challenges of extreme class imbalance and non-linear fraud dynamics in financial transaction data. By adopting a feature augmentation strategy that concatenates raw variables with autoencoder bottleneck representations and reconstruction errors, the hybrid architecture preserves core data attributes while recovering critical latent structural signatures. Validated across 85,442 test instances from the Kaggle Credit Card Fraud Detection dataset, the model achieved a robust held-out performance with a precision of 0.9621, a recall of 0.8581, an F_1 -score of 0.9071, and a realistic PR-AUC of 0.6942. Although cross-validation metrics showed no statistically significant performance delta over the baseline Random Forest ($p > 0.05$), the reconstruction error emerged as the second most influential predictor (importance score: 1.20), and t-SNE projections confirmed a distinct geometric separation of fraudulent instances within the latent manifold. These empirical findings confirm that augmenting supervised ensembles with unsupervised manifold features delivers a highly interpretable, audit-compliant, and operationally viable solution for real-time financial fraud monitoring.

REFERENCES

- Alhabib A., A., Alasiri A., F., Alharbi M., B., Ahmad S., Eljialy A. E. M. (2023) Credit Card Fraud Detection using Random Forest and K-Nearest Neighbors (KNN) Algorithms. International Conference on Cognitive Computing and Cyber Physical Systems, 383-395. Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-7371-8_30
- Bin Sulaiman R., Schetinin V., Sant P. (2022) Review of Machine Learning Approach on Credit Card Fraud Detection. Human-Centric Intelligent Systems, 2(1) 55-68. <https://doi.org/10.1007/s44230-022-00004-0>
- Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P. (2002) SMOTE: Synthetic Minority Over-Sampling Technique. Journal of Artificial Intelligence Research, 16, 321-357. <https://doi.org/10.1613/jair.953>
- Chen J., Shen Y., and Ali R. (2018) Credit Card Fraud Detection using Sparse Autoencoder and Generative Adversarial Network. 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 1054-1059. <https://doi.org/10.1109/IEMCON.2018.8614815>
- Chougule S., Mane K. (2026) Credit Card Fraud Detection using Machine Learning: A Comprehensive Survey and Analysis. Recent Advances in Computational Methods in Science and Technology, 421-430.
- El Alami M., Innan N., Shafique M., Bennai M. (2026) Comparative Performance Analysis of Quantum Machine Learning Architectures for Credit Card Fraud Detection. Applied Intelligence, 56(3), 83. <https://doi.org/10.1007/s10489-026-07110-7>
- Ezhilmathi S., Samy S. S. (2026) Fraud Activity Detection in Bitcoin Transaction using Adaptive Stacked Gated Recurrent Unit with Attention Mechanism Framework. Computational Economics, 68, 2661-2697. <https://doi.org/10.1007/s10614-025-11111-y>

- He H., Garcia E. A. (2009) Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284. <https://doi.org/10.1109/TKDE.2008.239>
- Kar S., Dasgupta S. (2024) Strategies to Mitigate Financial Fraud through Intellectual Capital Management. *Journal of Policy Options*, 7(4), 38-47. <https://resdojournals.com/index.php/jpo/article/view/388>
- Lin T-H., Jiang J-R. (2021) Credit Card Fraud Detection with Autoencoder and Probabilistic Random Forest. *Mathematics*, 9(21), 2683. <https://doi.org/10.3390/math9212683>
- Mallik A., Ahsan A., Shahadat M., Tsou J. (2019) Man-in-the-Middle-Attack: Understanding in Simple Words. *International Journal of Data and Network Science*, 3(2), 77-92. <https://doi.org/10.5267/j.ijdns.2019.1.001>
- Sun W., Qi Z., Shen Q. (2025) High-Recall Deep Learning: A Gated Recurrent Unit Approach to Bank Account Fraud Detection on Imbalanced Data. *2025 5th International Conference on Digital Society and Intelligent Systems (DSInS)*, 207-212. <https://doi.org/10.1109/DSInS68311.2025.11329905>
- Utami W., Nugroho L., Mappanyuki R., and Yelvionita V. (2020) Early Warning Fraud Determinants in Banking Industries. *Asian Economic and Financial Review*, 10(6), 604-627. <https://doi.org/10.18488/journal.aefr.2020.106.604.627>
- Vadapalli V., Ponugupati P., Dayam V., Arshad H. (2026) Credit Card Fraud Detection using Machine Learning. *Innovative Computing and Communications: Proceedings of ICICC 2025*, 8, 367-373. https://doi.org/10.1007/978-981-96-7526-5_40.
- Xie Y., Liu G., Yan C., Jiang C., Zhou M. (2023) Time-Aware Attention-Based Gated Network for Credit Card Fraud Detection by Extracting Transactional Behaviors. *IEEE Transactions on Computational Social Systems*, 10(3), 1004-1016. <https://doi.org/10.1109/TCSS.2022.3158318>